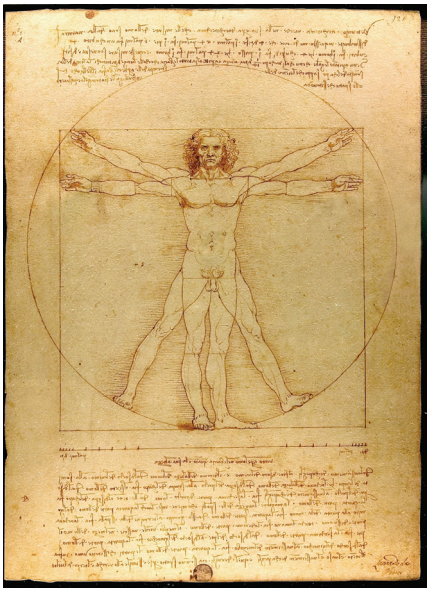
DISSERTATIONES SCHOLAE DOCTORALIS AD SANITATEM INVESTIGANDAM
UNIVERSITATIS HELSINKIENSIS

48/2017

PEDRO ALEXANDRE BENTO PEREIRA

The Human Microbiome in Parkinson's Disease and Primary Sclerosing Cholangitis



INSTITUTE OF BIOTECHNOLOGY AND
DIVISION OF GENETICS
DEPARTMENT OF BIOSCIENCES
FACULTY OF BIOLOGICAL AND ENVIRONMENTAL SCIENCES
DOCTORAL PROGRAMME IN INTEGRATIVE LIFE SCIENCE
UNIVERSITY OF HELSINKI

Institute of Biotechnology
and
Division of Genetics
Department of Biosciences
Faculty of Biological and Environmental Sciences
Doctoral School in Health Sciences (DSHealth)
Doctoral Programme in Integrative Life Science (ILS)
University of Helsinki

The Human Microbiome in Parkinson's Disease and Primary Sclerosing Cholangitis

Pedro Alexandre Bento Pereira

ACADEMIC DISSERTATION

To be presented for public examination with the permission of the Faculty of Biological and Environmental Sciences of the University of Helsinki at Metsätieteiden talo (Latokartanonkaari 9; accessible route Latokartanonkaari 7), room 105 (B6), on the 22nd of September 2017, at 12 o'clock noon.

Helsinki 2017

Supervisor	Research Director & Docent Petri Auvinen Institute of Biotechnology University of Helsinki, Finland
Thesis advisory committee	Professor Johanna Björkroth Department of Food Hygiene and Environmental Health Faculty of Veterinary Medicine University of Helsinki, Finland
	Professor Liisa Holm Institute of Biotechnology and Department of Biosciences Faculty of Biological and Environmental Sciences University of Helsinki, Finland
Pre-examiners	Docent Anne Salonen Department of Bacteriology and Immunology Faculty of Medicine University of Helsinki, Finland
	Professor Lars Engstrand Department of Microbiology, Tumor, and Cell Biology Karolinska Institute, Sweden
Opponent	Docent Reetta Satokari Immunobiology Research Program Faculty of Medicine University of Helsinki, Finland
Custos	Professor Liisa Holm Institute of Biotechnology and Department of Biosciences Faculty of Biological and Environmental Sciences University of Helsinki, Finland

Layout: Tinde Päivärinta
ISBN 978-951-51-3607-7 (paperback)
ISBN 978-951-51-3608-4 (PDF; http://ethesis.helsinki.fi)
ISSN 2342-3161 (print)
ISSN 2342-317X (online)
Hansaprint, Vantaa, 2017

Cover image by Leonardo da Vinci (Vitruvian Man). Photo by Luc Viatour / https://Lucnix.be

In memory of my father

CONTENTS

LIST OF ORIGINAL ARTICLES

I. Scheperjans F, Aho V, **Pereira P**, Koskinen K, Paulin L, Pekkonen E, Haapaniemi E, Kaakkola S, Eerola-Rautio J, Phoja M, Kinnunen E, Murros K, Auvinen P (2015) Gut microbiota are related to Parkinson's disease and clinical phenotype. Movement Disorders, 30:350-358

II. **Pereira P***, Aho V*, Paulin L, Pekkonen E, Auvinen P, Scheperjans F (2017) Oral and nasal microbiota in Parkinson's disease. Parkinsonism and Related Disorders, 38:61-67
** Equal contributors*

III. **Pereira P**, Aho V, Arola J, Boyd S, Jokelainen K, Paulin L, Auvinen P, Färkkilä M. Bile microbiota in primary sclerosing cholangitis: effects on disease progression and development of biliary dysplasia.
The draft version contained in this thesis is the exact version submitted to pre-examination. A modified version has now been published on PLOS ONE journal.

Figure 3 of this dissertation is based on unpublished data.

AUTHOR'S CONTRIBUTION

I. PP extracted DNA from the stool samples, made the DNA amplicon libraries for sequencing, carried out bioinformatic processing of the raw sequence data, contributed to the interpretation of the statistical analysis, and contributed to the writing of the manuscript.

II. PP carried out the bioinformatic processing of the raw oral sequence data, designed the statistical analysis of the oral data, carried out the statistical analysis of the oral data, contributed to the interpretation of the nasal results, wrote the section on oral results, and contributed to the writing of the overall manuscript.

III. PP extracted DNA from the bile samples, made the DNA amplicon libraries for sequencing, carried out the bioinformatic processing of the raw sequence data, did most of the design of the statistical analysis, carried out most of the statistical analysis, contributed to the interpretation of the results, and wrote most of the manuscript.

ABSTRACT

Parkinson's disease (PD) and Primary Sclerosing Cholangitis (PSC) are idiopathic diseases for which an external agent, e.g. a microorganism, has been hypothesised as being causally related to the condition or as having a potential role in disease progression. In the present work, we present evidence that gut and oral microbiota, but not nasal, differ between controls and PD patients, and that, regardless of hypothetical causal relationships, some of these changes show potential for use as biomarkers for diagnostic purposes. Although none of our studies are designed to investigate causality, we nevertheless find no suggestion that the bile microbiota is aetiologically related to PSC. On the other hand, our results suggest that *Streptococcus* may contribute to disease progression. All the studies are unified by the same methodology, consisting on case-control studies using high-throughput amplicon sequencing of the bacterial 16S rRNA gene, followed by bioinformatic processing of the raw data, and finally by statistical analysis.

ABSTRAKTI

Parkinsonin tauti (PD) ja primaarinen sklerosoiva kolangiitti (PSC) ovat idiopaattisia sairauksia joiden synnyn tai etenemisen on epäilty liittyvän ulkopuoliseen tekijään, esimerkiksi johonkin mikro-organismiin. Tässä tutkielmassa esittämämme löydökset osoittavat, että suoliston ja suun mikrobit poikkeavat Parkinsonin tautia sairastavien potilaiden sekä kontrollihenkilöiden välillä, mutta nenän mikrobit eivät. Riippumatta siitä, liittyvätkö löydetyt erot taudin etenemiseen, osa niistä voisi mahdollisesti toimia biomarkkereina tautidiagnostiikassa. Vaikka tutkimuksiamme ei ole varsinaisesti suunniteltu selvittämään syy-seuraussuhteita, emme löydä viitteitä siitä, että sapen mikrobit vaikuttaisivat PSC:n syntyyn. Toisaalta tulostemme perusteella näyttäisi siltä, että *Streptococcus* saattaa edistää taudin pahenemista. Kaikkia tutkimuksia yhdistää sama metodologia: ne ovat tapausverrokkitutkimuksia, jotka perustuvat bakteerien 16S rRNA:ta koodaavan geenin suuren kapasiteetin amplikonisekvensointiin jota seuraa sekvenssitiedon bioinformaattinen käsittely ja lopuksi tilastollinen analyysi.

Kindly translated to Finnish by Velma Aho.

ABBREVIATIONS

BMAA	β-N-methylamino-L-alanine
CNS	central nervous system
ENS	enteric nervous system
GABA	gamma-aminobutyric acid
GLM	generalised linear model
IBD	inflammatory bowel disease
LPS	lipopolysaccharide
OTU	operational taxonomic unit
PCR	polymerase chain reaction
PD	Parkinson's disease
PNS	peripheral nervous system
PSC	primary sclerosing cholangitis
rRNA	ribosomal ribonucleic acid
SCFA	short-chain fatty acid
SIBO	small intestinal bacterial overgrowth

1. INTRODUCTION

1.1. A Brief Overview of the Human Microbiome in Health & Disease

The human microbiome refers to the microbial communities found in and on the human body. These communities are composed of bacteria, archaea, and microbial eukaryotes, as well as human and microbial viruses (Aho *et al.* 2015). The term *microbiome* was coined by Joshua Lederberg in 2001 in the context of human microbial communities, and was defined as "the ecological community of commensal, symbiotic, and pathogenic microorganisms that literally share our body space and have been all but ignored as determinants of health and disease" (Lederberg & McCray 2001). The term was later adopted by microbial ecologists as referring to microbial communities found in and on any kind of environment, and the term also took a more inclusive meaning by taking into account the physical environment itself within the definition, thus better reflecting the *biome* concept in theoretical ecology (Aho *et al.* 2015). The term *microbiota* has a similar meaning but ignores the physical environment, thus referring strictly to the collection of organisms in a given environment. These two terms are often used as synonyms both colloquially and in the literature but in practice the intended meaning is clearly understood from context. Much confusion remains in the literature regarding the terms metagenomics and metagenome. A *metagenome* is the collection of all genetic material, ideally assembled into complete genomes, found in a given environment. In practice, however, obtaining complete genomes from most microbial environments is still technically problematic except by using very deep sequencing in the most extreme, low-energy, and/or highly oligotrophic environments. For the most part, the term is used specifically in the context of microbial ecology but conceptually can refer also to macro-organisms (e.g. Bourlat *et al.* 2013). *Metagenomics*, on the other hand, is a body of techniques for obtaining information on metagenomes, of which *shotgun metagenomics* is an example. As in genomics, metagenomics has the advantage, unlike 16S rRNA gene surveys, of providing direct evidence for the genetic make-up of microbial communities. This means that it is possible to infer functional information about the organisms, for example the distribution of metabolic pathways in the community. However, metagenomics does not provide information on gene expression, and thus we can only obtain information on the functional potential of the community. A complementary approach is *metatranscriptomics*, which, in an analogous way to *transcriptomics*, can provide data on the activities of the communities at the time the samples are taken. Metatranscriptomics relies on RNA sequencing and analysis of gene expression. The studies presented in this thesis (I, II, III) are, strictly speaking, not metagenomics-based. Instead, they are better labelled as *phylogenetic marker gene surveys* using high-throughput 16S rRNA gene amplicon sequencing technology. Thus, no direct information on the functional capabilities of the communities can be obtained, except indirectly through taxonomic identification of the sequences followed by extrapolation.

Estimations of the quantity of microorganisms present in the human body at any given time has led to the suggestion that the number of bacterial cells, specifically, may exceed our own by an order of magnitude (Savage 1977, Turnbaugh *et al.* 2007). These estimations have been criticized on the basis that microbiome-related interpersonal variation is very high and that estimations of the number of human cells themselves are surprisingly unreliable, ranging from $5x10^{12}$ to $7x10^{16}$ cells, depending on the author (Bianconi *et al.* 2013, Torres 2014). The number of different protein-coding genes present in these communities, even by conservative estimates, far exceeds the estimated number of protein-coding genes in humans: ~3.3 to ~9.9 million bacterial genes *vs* ~19.000 human genes (Qin *et al.* 2010, Ezkurdia *et al.* 2014, Li *et al.* 2014). The ecological relationships of these microbiota with the human host vary between commensalism,

mutualism, amensalism, and parasitism, with any given microorganism potentially being able to shift its relationship with its host depending on present conditions and opportunities. For example, opportunistic pathogens represent a type of amphibiosis (Rosebury 1962), and consist of microorganisms that are commonly present in healthy individuals in a state of non-pathogenicity, usually in low abundances, and that may become pathogenic due to e.g. abnormal growth promoted by microbial community disruption, access to normally inaccessible tissues, a compromised immune system, etc. It is still common to refer to the non-pathogenic microbiota simply as commensals in the literature, but this is an ecological oversimplification.

Ilya Metchnikoff, whose work laid the foundation for the modern field of immunology together with Paul Erlich (Kaufmann 2008), was the first to actively promote the essential role of microbiota in the general health of the human host (Metchnikoff 1908, Bhattacharjee & Lukiw 2013), thus joining the ranks of other contemporary visionaries such as Vladimir Vernadsky and Sergei Winogradsky in his holistic approach to microbiology. Since then, a voluminous body of work has been produced on the relationship between the human microbiome and health & disease in general (see e.g. Shreiner *et al.* 2015, Lynch & Pedersen 2016, Young 2017 for an overview). For an historical perspective on microbial endocrinology and the microbiota-gut-brain axis in particular, see the extensive three-part literature review by Bested *et al.* (2013a), Bested *et al.* (2013b), and Bested *et al.* (2013c), and for an introduction to the subject see e.g. Lyte & Cryan (2014) and Dinan & Cryan (2017b).

The introduction of high-throughput DNA sequencing technology has enabled culture-independent methods to make major strides in the detection and characterization of bacteria and other microorganisms, including in locations of the human body that have previously been considered sterile environments except under unhealthy states, e.g. in the lower respiratory tract of healthy individuals (Morris *et al.* 2013, Segal *et al.* 2013). There are many challenges in 16S rRNA gene survey studies, in particular in relation to the inability to differentiate between free DNA and living cells, the low biomass of the starting material, and possible contamination either during sampling or due to contaminated reagents for sample processing (Pezzulo *et al.* 2013, Lazarevic *et al.* 2014, Salter *et al.* 2014, Aho *et al.* 2015). To a greater or lesser degree, these are issues that have the potential to affect any 16S rRNA gene-based study, and need to be taken into account during the planning of sampling methodology, DNA extraction and library construction, and sequence data processing.

Pyrosequencing was commercially developed in its most well-known incarnation by the company 454 Life Sciences, founded in 2000 by Jonathan Rothberg. Released in 2005, the original instrument's name was Genome Sequencer GS20, and it catapulted high-throughput (pyro)sequencing into the limelight. The company was eventually bought by Roche Diagnostics in 2007, and by mid 2016 Roche discontinued the 454 platform. It was during 2012-2013 that we used the 454 platform to sequence the stool samples that would provide the data for Study I. 454-pyrosequencing used a sequencing-by-synthesis (SBS) approach, meaning that sequencing of DNA relied on the enzymatic activity of a DNA polymerase. SBS is still by a large margin the most common technology for DNA sequencing. The 454 platform used emulsion polymerase chain reaction (PCR) for amplification of DNA fragments on the surface of beads. These would then be moved to a well (acting as a micro-reactor) on the surface of a picotiter plate, where the actual SBS process would occur. Incorporation of a nucleotide would release a pyrophosphate that would be converted to ATP. Luciferase would then use the released ATP to oxidize luciferin, thus producing a light signal detectable by a camera. In its halcyon days, 454 technology would allow read lengths of circa 800-1000 base-pairs with FLX+ chemistry or 400-500 base-pairs with Titanium chemistry. We used the latter for amplicons in Study I. 454 would produce circa

one million sequences per run with both chemistries, making it a veritable workhorse for 16S rRNA gene amplicon sequencing studies in microbial ecology (Margulies et al. 2005, Rothberg & Leamon 2008, Goodwin et al. 2016). The 454 platform was eventually supplanted, within the context of microbial ecology, by the MiSeq platform from Illumina. This company was founded in 1998 by David Walt, and it acquired the company Solexa in 2007, which had also been founded in 1998. Solexa released its Genome Analyser DNA sequencing platform in 2006, which was then further optimised and re-released by Illumina in 2009. The MiSeq platform itself was released in 2011 and, like 454, it's also based on SBS technology. It uses bridge PCR amplification of DNA fragments anchored to the surface of a flow-cell, forming clone clusters or "colonies", and detection of fluorescence for determination of nucleotide incorporation. We used MiSeq in 2014-2015 to sequence the samples for studies II and III, since our laboratory had by now acquired the platform. The MiSeq platform in its present form allows for the sequencing of circa 25 million reads per run (50 million paired-end), with maximum read lengths up to 300 + 300 base-pairs pair-ended with v3 chemistry. It's read lengths are shorter than those from 454, but this can be compensated by using paired-end reads (sequencing matched forward and reverse complements) if sequencing longer amplicons is desired (e.g. V1-V3 regions of the 16S gene, as in the studies presented in this thesis; Goodwin et al. 2016). The main advantage of the MiSeq platform over the Roche 454 is that it produces substantially more reads per run. This enables samples to be sequenced more deeply, and thus provide a more representative view of community composition, including rarer taxa.

A growing number of studies are investigating the human microbiome not only in healthy individuals but also in relation to several medical conditions. A full survey of the literature is beyond the scope of the present introduction, but examples such as cystic fibrosis (Surette 2014, McGuigan & Callaghan 2015, Bacci *et al.* 2016, Feigelman *et al.* 2017), chronic obstructive pulmonary disease (Sze *et al.* 2014), asthma (Beigelman *et al.* 2014), autism (De Angelis *et al.* 2013, Kang *et al.* 2013, Krajmalnik-Brown *et al.* 2015, Son *et al.* 2015), Alzheimer's disease (Harach *et al.* 2017), Parkinson's disease (Hasegawa *et al.* 2015, Keshavarzian *et al.* 2015, Unger *et al.* 2016, Hill-Burns *et al.* 2017), arthritis (Scher *et al.* 2013, Chen *et al.* 2016a), colon cancer (Ou *et al.* 2013), diabetes (Kostic *et al.* 2015), and primary sclerosing cholangitis (Kummen *et al.* 2016, Sabino *et al.* 2016, Torres *et al.* 2016), illustrate well the importance that human microbiome studies have achieved in medical science during the last decade. It is in the context of this body of work that the three studies presented in this thesis were conducted.

1.2. The Rationale for Microbiome Studies in Parkinson's Disease

Parkinsons's disease (PD) is an idiopathic condition that manifests itself mostly through impaired motor function, gastrointestinal, olfactory, and oral disturbances, as well as behavioural problems (Sveinbjornsdottir 2016). The early classical symptoms consist of muscular rigidity, slow movements, resting tremors, changes in gait, and postural instability. Later symptoms include abnormal behaviour, depression, and dementia. However, there is a range of symptoms that may manifest at an earlier stage, in many cases preceding formal diagnosis of the disease by ~10 years (Schrag *et al.* 2015). Chiefly among these is gastrointestinal dysfunction, which is present in 80% of the cases and is mostly represented by slow gastrointestinal transit (constipation), worsening as the disease progresses. The latter condition may also predispose patients to small intestinal bacterial overgrowth (SIBO; Fasano *et al.* 2013, Niu *et al.* 2016).

PD is a progressive neurodegenerative disorder associated with aging, and is the second most common disease in this class at the global level (de Lau & Breteler 2006, Pringsheim *et al.*

2014). There are genetic factors involved in PD, with 28 distinct chromosomal regions possibly being related to the disease. In these, six genes have been identified containing mutations that are consensually thought to independently cause monogenic PD, the non-idiopathic version of the disease, although these six gene variants altogether only explain 3-5% of PD occurrence (Klein & Westenberger 2012). Also, epidemiological studies using monozygotic and dizygotic twins suggest that in late-onset PD environmental factors are more important than in early-onset PD, where genetic factors seem to be more relevant (Wirdefeldt *et al.* 2011). Other epidemiological evidence has linked PD with various exogenous factors such as pesticides, metals, dietary habits, and occupational history, of which only pesticides and intake of dairy products show some consistency of results (see Wirdefeldt *et al.* 2011 for a comprehensive review).

The disorder strongly affects a region of the brain called *substantia nigra*, which is involved in motor control and coordination, by severely impairing and killing dopaminergic neurons. Dopaminergic transmission to the striatum is important in the regulation of voluntary and involuntary muscle movements. A classic pathologic feature of the PD-affected brain is the presence of Lewy bodies, which consist of inclusions of accumulated alpha-synuclein fibrils. Lewy bodies are believed to be aggresomes, inclusions of misfolded proteins derived from the inability of cells to degrade them. Alpha-synuclein is abundant in the presynaptic axon terminals. Its function is not well understood, but it is thought to be involved in maintaining the supply of synaptic vesicles at the axon terminal's active zone for release (Diao *et al.* 2013).

Braak and colleagues (Braak *et al.* 2003a, Braak *et al.* 2004) have suggested the following sequence of events for the brain pathology of PD:

a) before the manifestation of PD-related motor symptoms, Lewy bodies appear in the dorsal motor nucleus of the vagus nerve and the pathology spreads from there to the *medulla oblongata/pontine tegmentum* (which are associated with breathing, heart rate, and blood pressure control; and with sensory and motor functions, control of the stages of sleep, respiratory rate, and arousal and vigilance, respectively). The olfactory bulb/anterior olfactory nucleus is also affected at this stage, followed somewhat later and independently by the lower raphe nuclei;

b) as the disease progresses, it spreads to the *substantia nigra* and adjacent areas of the midbrain and forebrain, and possibly the first classical symptoms (i.e. motor) appear;

c) finally, the pathology spreads throughout the neocortex, which has functions related to sensory perception, motor control, and language, among others.

This sequence of events has become known as the Braak Hypothesis, and given its association with the vagus nerve, it led to the suggestion that PD could be aetiologically related to an environmental agent (a bacterium, a virus, or a toxin) entering the human body through the nose and then from there to the gut, subsequently gaining access to the central nervous system (CNS) after breaching the gut mucosal barrier, and from there spreading into the brain (Braak *et al.* 2003b). In fact, it is known that PD pathology (i.e. Lewy bodies) affects not only the CNS but also the peripheral nervous system (PNS) and particularly the enteric nervous system (ENS), including in putative pre-symptomatic cases (Wakabayashi *et al.* 1988, Braak *et al.* 2006, Dickson *et al.* 2008, Beach *et al.* 2010, Lebouvier *et al.* 2010, Shannon *et al.* 2012, Visanji *et al.* 2013). The environmental agent would therefore gain entry to the neural axons (nerve fibers that conduct electric impulses) of the ENS through the enteric epithelium and be transported trans-

synaptically to the CNS via the vagus nerve into the medulla. Regarding entry into neurons, the uptake of substances from outside the neuronal space at the axon terminals is known to occur, with several neuroactive substances commonly gaining entry in this way (e.g. through receptor-mediated endocytosis; Braak *et al.* 2003b). This could provide the possible route of entry into the nervous system for a putative pathogenic agent. Also, olfactory and taste dysfunction is common in early PD, as well as early Lewy pathology of the olfactory bulb and oral raphe nuclei, although this would not be the preferred route of pathology spread into the brain. Instead, the pathology would essentially be restricted to the olfactory and oral areas, given that the observed pathology doesn't seem to spread substantially to the brain from these peripheral areas (Braak *et al.* 2003b). This led to the suggestion that the nasal route could be involved in tandem with a gastrointestinal origin, after swallowing nasal secretions containing the putative agent. The oral route would probably not be the main point of origin of the disease because the pathology seems to manifest at this location later than in the nasal structures, and also because direct neuronal connections exist only between the main olfactory areas and the *substantia nigra* (Hawkes *et al.* 2007, Lerner & Bagic 2008, Sengoku *et al.* 2008, Del Tredici & Braak 2016). It's interesting to note in this context that there is experimental evidence for the entry of neurotropic viruses into the brain through the nasal route (Hawkes *et al.* 2007). On the other hand, Beach *et al.* (2010) showed support for greater involvement of the oral areas, with their study showing alpha-synuclein pathology of the submandibular glands in PD, which are salivary glands and could be related to the decreased saliva production in PD. In fact, within the gastrointestinal tract, the submandibular glands and the lower oesophagus were the most affected areas, followed by the stomach, small bowel, large bowel, and rectum, in order of decreasing pathology severity. No pathology was found in the upper oesophagus.

Evidence has been presented that alpha-synuclein could behave as a prion, essentially being the driver of the pathology (Li *et al.* 2008, Visanji *et al.* 2013, Kordower 2014, Chauhan & Jeans 2015, Chu & Kordower 2015). Misfolded proteins, including alpha-synuclein in PD, could be associated with various neurodegenerative disorders, including via cell to cell spread from the ENS to the CNS (Lerner & Bagic 2008, Aguzzi & Rajendran 2009, Brundin *et al.* 2010, Natale *et al.* 2011a, Natale *et al.* 2011b, Polymenidou & Cleveland 2011, Prusiner 2012, Olanow & Brundin 2013). A number of lab experiments, both *in vitro* and *in vivo*, have also been conducted regarding this hypothesis (see previous references in this paragraph for reviews), with one of the most interesting involving rats injected with misfolded alpha-synuclein that showed that amyloid pathology can spread using axonal transport in a prion-like form from the gut to the brain (Holmqvist *et al.* 2014). Another study showed that induced overexpression of human alpha-synuclein in the brain of rats resulted in accumulation in the vagal terminals of stomach walls (Ulusoy *et al.* 2017). In the first study the route of spread was confirmed to be the vagus nerve, and this was very probably the case also in the second study. These results reflect the issue that detection of pathology at a particular location does not by itself translate into place of initiation of disease spread. In fact, Beach *et al.* (2010) have suggested, based on their study results, the reversed order of events, with alpha-synuclein pathology possibly spreading from both the olfactory regions and the brain to the gut. Del Tredici and Braak (2012) also produced evidence in support of this hypothesis, showing that the spinal cord may get affected only after the brain.

Regardless of place of pathology initiation and spread direction, it is also possible that the misfolded alpha-synuclein is somehow related to a microbial agent. For example, β-N-methylamino-L-alanine (BMAA) has been found to be elevated in parkinsonism-dementia complex of Guam (Banack *et al.* 2010, Bhattacharjee & Lukiw 2013). BMAA is an oxidative

stress-inducing neurotoxin in the form of an amino acid that is speculated to be produced by *Melainabacteria*, a group of flagellated, obligate-fermenting, molecular hydrogen producing, B and K vitamin synthesizing, non-photosynthetic cyanobacteria found in the human gut, which were originally classified as their own phylum but are now included in the phylum *Cyanobacteria* (Di Rienzi *et al.* 2013, Soo *et al.* 2014, Soo *et al.* 2017). More interestingly, BMAA has been linked to intra-neuronal protein misfolding and neurodegeneration. Also, a recent study using mice models for Alzheimer's disease showed that the diseased mice's gut microbiota not only differed from healthy control mice, but that diseased gnotobiotic mice had substantially lower amounts of Aβ-amyloid plaque in their brains than the diseased non-gnotobiotic mice. By transferring gut microbiota from non-gnotobiotic diseased and non-diseased mice into "healthy" gnotobiotic mice, the researchers showed that the latter developed significantly more amyloid plaques if the transfer was made from the diseased mice (Harach *et al.* 2017). It is also interesting to note in this context that the gut microbiota widely produces functional amyloid peptides for biofilm formation, cell aggregation, host-cell adhesion, toxins, and protection against host defences (Larsen *et al.* 2007, Blanco *et al.* 2012, Garcia *et al.* 2013, Hufnagel *et al.* 2013, Mulligan & Chakrabartty 2013, Schwartz & Boles 2013, Hill & Lukiw 2015). For example, *Escherichia coli*, as well as others in the *Enterobacteriaceae* family, naturally produces a functional amyloid, curli, from at least six proteins encoded by two operons (Barnhart & Chapman 2006). The major peptide structural subunit, CsgA, contains amyloidogenic repeat motifs shared by human prions and by alpha-synuclein (Chen *et al.* 2016b). *Cross–seeding*, in which one amyloidogenic protein initiates aggregation of a different amyloid protein, has been demonstrated experimentally (Lundmark *et al.* 2005). Chen *et al.* (2016b) showed evidence that cross-seeding between curli and alpha-synuclein occurs by exposing aged Fischer 344 rats and transgenic nematodes (*C. elegans* expressing human alpha-synuclein) to wild-type curli-producing *E. coli*, with the control rats and nematodes being exposed to engineered *E. coli* lacking the curli-production operons. The aged rats, orally exposed to the bacteria, showed enhanced alpha-synuclein aggregation in the brain and deposition in the gut, while the nematodes, who naturally feed on *E. coli*, showed increased alpha-synuclein aggregation starting in the head and then spreading to the tail.

Abnormal microbial community composition and structure could also conceivably lead to neurological disease, as opposed to a particular microbial agent. For example, a significant number of neuroendocrine hormones are produced by bacterial communities in the animal gut, many of them sharing the same protein structure and biosynthetic pathways found in mammal lineages, e.g. members of the catecholamine family, like tyrosine and epinephrine (adrenaline), complete pathways of which can be found in *Escherichia coli* and other bacteria, acetylcholine, histamine, serotonin, agmatine, GABA, and various others. More importantly, some of these have been found in amounts theoretically capable of eliciting physiological responses in the animal host (Lyte & Cryan 2014). Thus, bacteria as well as other microorganisms (e.g. yeasts) are capable of producing, as well as responding to, the same neurohormones utilized by the host, since many possess receptors for these neuroactive substances, in some cases with high-affinity such as in *Pseudomonas* (Guthrie *et al.* 2000). It is not, however, clear what the function of these neuroactive chemicals are in microorganisms, with a few exceptions like e.g. GABA, which seems to be involved in resistance to acidic pH in some *Lactobacillus* species and in the germination of *Bacillus* spores (Foerster & Foerster 1973, Su *et al.* 2011). It is quite possible that these substances are used as chemical signals in quorum-sensing. This opens the possibility that microbe-host bidirectional signalling occurs. For a comprehensive review of these matters see Lyte & Cryan (2014) and references therein. Thus, our two pilot studies on PD (I & II) aimed at obtaining clues

from the colonic, oral, and nasal bacterial communities about possible candidates for bacterial infectious agents in accordance with the Braak Hypothesis, as well as insight into possible overall community dysfunction.

1.3. The Rationale for Microbiome Studies in Primary Sclerosing Cholangitis

Primary Sclerosing Cholangitis (PSC), like Parkinson's, is an idiopathic disease. In PSC, chronic inflammation of the liver leads to strictures of the bile ducts and eventually to cholestasis and secondary biliary cirrhosis (Hirschfield *et al.* 2013). The inflammation is associated with a substantially increased risk of development of biliary dysplasia and cholangiocarcinoma (Bergquist *et al.* 2002). PSC results from obstruction of the bile ducts and biliary epithelial tissue damage, mediated by inflammation, leading to decreased bile flow and accumulation of toxic bile products. Together with Primary Biliary Cirrhosis (PBC), they represent the major clinical pathologies in chronic cholestatic liver disease (Mattner 2016).

Liver produces bile, which flows through the hepatic ducts into the gallbladder, where it is stored and concentrated mostly by removal of water, and from which it is then released into the duodenum through the common bile duct. Bile consists mostly of water and bile acids. These acids serve several important functions, among which is the inhibition of bacterial growth and adhesion for protection against bacteria ascending from the small intestine through the biliary tract. For this purpose, bile is combined with secreted immunoglobulin A (IgA; Wu *et al.* 2004, Mattner 2016). The formation of mucins by the biliary epithelial cells to form a mucus layer, the presence of bacterial cell wall breakdown products including lipopolysaccharides (LPS) and lipoteichoic acid in human bile, as well as the local expression of immune receptors and production of antimicrobial peptides in the biliary tract, all indicate that the bile tract is prepared for the presence of bacteria and the risk of infection (Sasatomi *et al.* 1998, Tsuneyama *et al.* 2001, Verdier *et al.* 2015).

Obstruction of the bile flow in PSC not only damages the integrity of the epithelial layers of the bile ducts but may also leads to alterations in putative microbial bile communities (e.g. in the gallbladder) and to an increase in the susceptibility to infection (Miyake & Yamamoto 2013, Verdier *et al.* 2015). For example, it is known that a number of microorganisms are resistant to bile acids, such as *Escherichia coli* and *Helicobacter* spp. (Brook 1989, Fox *et al.* 1995, Carpenter 1998, Ganzle *et al.* 1999, Hirai 1999). Nevertheless, a healthy biliary tract has until very recently been considered to be a sterile one (Verdier *et al.* 2015). In general, putative biliary microbial communities appear to be similar to those of the upper digestive tract, and have been suggested to originate from the small intestine (Shen *et al.* 2015, Ye *et al.* 2016), although there is no agreement on what a "healthy" microbial community, if any exists, should be like (Verdier *et al.* 2015).

It has been suggested that microorganisms, microbial metabolites or derivatives could be associated with the etiopathogenesis of PSC (Tabibian *et al.* 2013, Eksteen 2014, Tabibian *et al.* 2014, Mattner 2016). There have also been a number of studies that provide indirect evidence for the involvement of microorganisms in the aetiology of PSC and/or in the disease's progress. For example, decreased bile acid secretion has been linked to SIBO in liver cirrhosis (Bauer *et al.* 2001). There is also a close association between PSC and inflammatory bowel disease (IBD), which is thought to be linked to dysbiosis of the gut (Loftus *et al.* 1997, Packey & Sartor 2009, Reiff & Kelly 2010, Eaton *et al.* 2013, Karlsen & Boberg 2013). Bacteria have also been cultured frequently from the bile of patients with acute cholangitis or cholecystitis, and most are found

also in the intestinal tract, e.g. *Escherichia coli*, *Klebsiella*, *Enterobacter*, *Pseudomonas*, and *Citrobacter* spp. (Csendes *et al.* 1996). More importantly, their detection has been associated with disease progression to severe cholangitis and increased mortality rates. An experiment that consisted in inducing SIBO in the small intestine of rats triggered hepatobiliary inflammation in a way that resembled the histological and cholangiographic features of PSC (Lichtman *et al.* 1991). Furthermore, administration of an enzyme for peptidoglycan degradation, mutanolysin, to those rats improved the observed hepatobiliary pathology and also reduced inflammation (Lichtman *et al.* 1992).

Evidence of a more direct nature exists for a link between bacteria and PSC in humans. Starting with colonic bacteria, a study based on ileocecal biopsies and using phylogenetic microarrays found that, compared to ulcerative colitis (UC) and controls, colonic mucosa-associated microbiota in PSC showed lower diversity and a reduced abundance of uncultured *Clostridiales II* (Rossen *et al.* 2015). Some studies of gut communities have also been performed with DNA sequencing-based methods, with one study based on biopsies of the terminal ileum and the large intestine finding that the *Barnesiellaceae* family, the genus *Blautia*, as well as a number of OTUs (Operational Taxonomic Units) mostly from the order *Clostridiales* were enriched in PSC (Torres *et al.* 2016). Finally, two studies using stool samples found increased abundances for the genera *Enterococcus*, *Fusobacterium*, and *Lactobacillus* (Sabino *et al.* 2016), and *Veillonella* (Kummen *et al.* 2016) when comparing healthy controls to PSC patients, with both studies also finding lower bacterial diversity in PSC.

Finally, bile-based studies in PSC have also been performed, with an early study showing that diverse bacteria, particularly alpha-haemolytic streptococci, could be cultured from bile samples of patients with PSC, but not from those with primary biliary cirrhosis, thus also suggesting that contamination was not an issue in their studies since the methods of extraction used were the same (Olsson *et al.* 1998). They also suggested that previous endoscopic retrograde cholangiography (ERC) procedures were the likely source of the identified bacteria, ultimately originating from the small intestine. A more recent study used 16S rRNA gene amplicon sequencing to characterize the biliary bacterial communities of 39 PSC patients (Folseraas *et al.* 2012), but this study focused on possible links between microbial communities and human genetic features, and did not include any non-PSC controls. Thus, our study (III) represents a timely addition to this body of work on PSC, investigating the microbial communities in bile retrieved during ERC examinations, as well as the possible role for microbiota both in the aetiology of PSC and in the disease's progression.

1.4. The Role of Exploratory Studies in Human Microbiome Research

A criticism that is occasionally levelled against exploratory studies such as the ones presented in this thesis (I, II, III) is that they represent "hypothesis-free" research. This is a misunderstanding of the role these studies play in biomedicine in particular, and of exploratory research in general. Present scientific knowledge on microbiota-host interactions, on specific microorganisms, as well as on the potential role of the latter in a specific disease, is not developed to such an extent that would make it possible to formulate a specific scientific hypothesis amenable to translation into a clear-cut *statistical* hypothesis that could in turn be formally investigated, as is the case in e.g. experimental studies (Figure 1). Let us say that we want to test a certain scientific hypothesis, well-grounded on previous knowledge, that a particular microorganism should be overrepresented in a given population. In this case, we could design an observational (and confirmatory) study that would allow us to target specifically that organism and test our hypothesis. We wouldn't

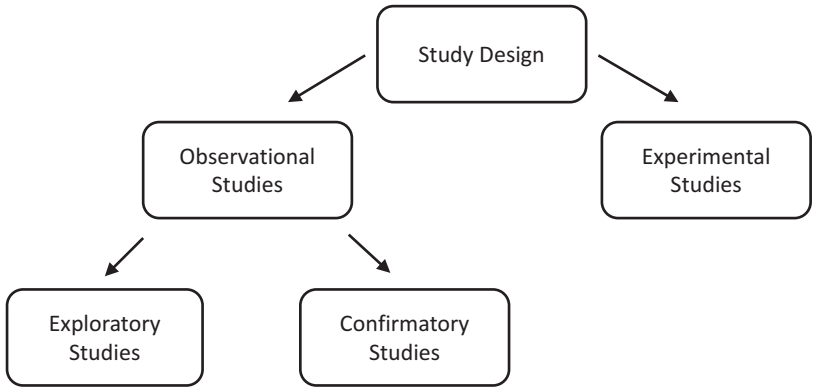


Figure 1. General overview of the relationship between observational and experimental studies. *More specific types of studies can fall into multiple categories. For example, case-control studies like the ones presented in this thesis can be used in either exploratory or confirmatory studies.*

need to use *p*-values either, especially under the Null Hypothesis Significance Testing (NHST) paradigm of Neyman and Pearson, since we could simply obtain point estimates for population abundance means and associated confidence intervals as measures of uncertainty (Cumming 2012, Rothman 2012).

Unfortunately, this is not possible in the kind of studies presented in this thesis. Our scientific hypotheses are of a more general character and there are no specific target organisms, as shown in the previous sections of the Introduction. Instead, what we want to know is if there is evidence for the involvement of *any* group of organisms in a disease, with differential abundance being a proxy for said potential involvement. To do this in practice, we need to use *p*-values under a NHST framework, or some other method that allows us to set a pre-specified threshold of interest to red-flag results from a screening procedure. The reason is that, given the immense number of taxa seen in human microbiome studies (hundreds to thousands per dataset), we need to filter the most promising results, and NHST provides a convenient and acceptable method to perform said filtering, if used cautiously. *p*-values can be verbally defined in the following way: assuming that the null hypothesis of zero effect is true, what is the probability of observing an effect at least as extreme as the one observed in this study, just by chance (Motulsky 2010)? In NHST this idea is extended further: above a certain alpha cut-off (usually set at 0.05) the results are considered as not providing enough evidence to reject the null hypothesis. Below the threshold, they are considered *statistically significant*. Essentially, NHST provides an *automated, dichotomous* answer, i.e. either there's a "difference" or there isn't, based on an arbitrary threshold under the form of the alpha (the *significance level* of the test). In practice, we need to acknowledge that using some kind of filter is desirable in the context of these studies, such that taxa whose differential abundances are statistically significant are *after* detection paid particular research attention.

A related criticism sometimes levelled against *p*-values in general and NHST in particular, e.g. Kruschke (2015), Cumming (2012), Gelman *et al.* (2014), is essentially that the null hypothesis is already known *a priori* to be false in most cases. There will (almost) always be some level of real variation, no matter how small the effect, associated with any comparison between populations, such that if we had unlimited access to samples we could always find a statistically significant difference. Therefore, the question is usually one of degree of difference, not if there is a difference. Although this is essentially correct, we nevertheless still need in practice to

use some form of filter to screen massive amounts of data, for the reasons argued above. For that purpose, the definition of a *p*-value has sometimes been colloquially reformulated into something more operational: assuming that the null hypothesis is, *for all practical purposes statistically indistinguishable from zero*, what is the probability of observing an effect at least as extreme as the one observed, just by chance? After obtaining the statistically significant results, one can then assess their effect sizes, data patterns, uncertainty estimates, etc, and ponder their potential biological relevance. If the effect is deemed biologically meaningless in a particular research context (e.g. very small but statistically significant effect size), then it can be ignored. Thus, exploratory studies are best seen as *hypothesis-generators*, since any clues derived from their results can then be used to conceive specific hypothesis amenable to investigation with targeted studies designed with that aim in mind. This means that our concern is not to demonstrate that there is definitely and beyond any reasonable doubt a connection between the microbiota and a given disease, but instead to evaluate if there is evidence for such a connection and if that evidence is substantial enough to warrant further investigation and/or to substantiate previously existing studies, especially if they used different methods from ours.

Finally, there is the question of sample size and statistical power (Figure 2). In experimental studies or in observational but confirmatory studies the researchers may have a clear idea of the effect sizes they deem interesting, as well as information on the expected standard deviations of the distributions of their variables of interest. However, this is usually not the case in exploratory studies such as the ones presented here. There are several reasons why this is the case. First, the effect that the explanatory variables may have on the dependent variable is usually not known or it is known only in broad qualitative terms. Second, the variation associated with the explanatory variable (in this case taxon abundance) is also not sufficiently constrained to allow meaningful

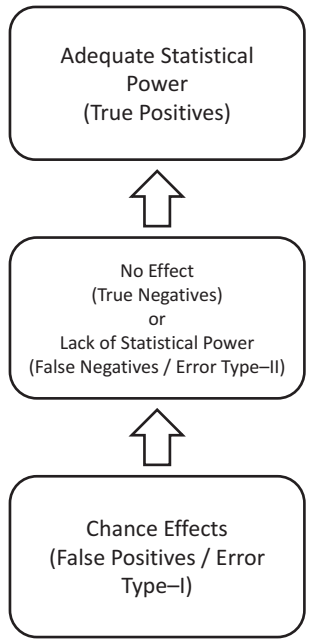


Figure 2. A simplified view of possible effects and their relationship to sample size. *Arrows denote increasing sample size. Notice that the figure is simplistic and for illustrative purposes only, e.g. chance effects can potentially occur even when the sample size is estimated as being "large" and adequate for detecting a hypothetical real effect.*

power calculation for a particular effect size. Third, we usually do not have enough biological knowledge to *a priori* determine what a meaningful effect size would be. And finally, we have to deal with hundreds to thousands of different taxa at the same time, and each of these will be a dependent variable in an associated statistical model. Therefore, in practice, power calculations for determination of sample size are usually not practical in the context of these studies. The practical alternative relies on a combination of the researchers' experience with data from human microbiome studies, background biological knowledge relevant to the problem at hand, and data from previously published studies that can potentially serve as a guideline (Debelius *et al.* 2016).

2. AIMS OF THE STUDIES

For the two Parkinson's disease studies (I & II), the objectives were:

1) to investigate the presence or absence of support for the hypothesis that PD could be aetiologically related to a putative pathogen invading the host from the gut, nasal, and/or oral areas. We specifically investigated the possibility of existence of a bacterial agent, although others are also conceivable (e.g. viral). A direct test of this hypothesis was not the aim here, since the studies were not designed to directly assess causality; instead, the studies were built as a first step in the assessment of the hypothesis.

2) We also aimed at evaluating the more general possibility that overall bacterial community composition and structure could be associated with PD, causally or not.

3) Finally, we also aimed at assessing the potential of microbiota to be used as biomarkers in PD.

To achieve these objectives, we used high-throughput 16S rRNA gene sequencing methods that allow the determination of the presence and relative amounts of most bacterial taxa in the most unbiased way possible.

As for Study III on bile microbiota and Primary Sclerosing Cholangitis, the aims were:

1) to compare bacterial abundances in controls to early disease stage PSC patients (ERC severity score <6) at their first ERC examination as a way to avoid any confounding effects derived from the ERC interventions, and to evaluate the evidence obtained (or lack of it) for the involvement of bacteria in the aetiology of the disease. Another statistical model was also used in which all early stage patients were included (to increase statistical power) while controlling for the potential effects of the number of ERC examinations directly through modelling.

2) To compare bacterial abundances during disease progression between early disease stage PSC patients and advanced disease stage patients (ERC severity score ≥ 6), while controlling for the number of ERC examinations through statistical modelling.

3) To compare bacterial abundances between the advanced disease stage and the dysplasia/carcinoma stage, while controlling for the number of ERC examinations through statistical modelling.

4) To assess bacterial differential abundances and their relationship to disease by using the ERC severity score directly, while controlling for the number of ERC examinations through statistical modelling.

5) And finally, to examine the impact of the number of ERC interventions on the microbiota, since previous reports suggested that cannulation during ERC procedures could be associated with infection of the bile ducts.

As in the case of studies I & II, we used high-throughput 16S rRNA gene sequencing methods that allow the determination of the presence and relative amounts of most bacterial taxa in the most unbiased way possible.

3. MATERIALS AND METHODS

3.1. DNA Library Preparation and Sequencing

3.1.1. Parkinson's Disease

Table 1 presents general clinical data for the study populations in Studies I and II. Only the main clinical variables are presented here. For more information consult the respective articles and supplementary material.

Table 1. Selected clinical data for Studies I & II.

Study	Number of Subjects	Selected Clinical Data % (C & PD)	Female Subjects %	Age (years, mean ± SD)
I	72 (C) *vs* 72 (PD) **PD Motor Phenotype (n/72):** TD: 23/72 MX: 9/72 PIGD: 40/72	**IBS (resolved):** 2.8% & 1.4% ***Wexner Score (median, IQR):** 2 [1-4] & 5 [3-9] ***Atrial Fibrillation:** 18.1% & 4.2% ***TIA or Ischemic Stroke:** 37.5% & 7.0% ***Levodopa:** 54.2% (PD only) ***COMT Inhibitor:** 15.3% (PD only) ***Dopamine Agonist:** 77.8% (PD only) ***MAO Inhibitor:** 70.8% (PD only) ***Anticholinergic:** 8.3% (PD only) ***Warfarin:** 15.3% & 1.4% ***Statin:** 54.2% & 20.8%	50.0 (C) & 48.6 (PD)	64.5 ± 6.9 (C) & 65.3 ± 5.5 (PD)
II	**Oral:** 76 (C) *vs* 72 (PD) **Nasal:** 67 (C) *vs* 69 (PD)	***Atrial Fibrillation:** **O:** 17.1% & 4.2%, **N:** 19.4% & 4.4% ***TIA or Ischemic Stroke:** **O:** 35.5% & 5.6%, **N:** 37.3% & 5.8% ***Levodopa (PD only):** **O:** 52.8%, **N:** 52.2% ***COMT Inhibitor (PD only):** **O:** 13.9%, **N:** 16.0% ***Dopamine Agonist (PD only):** **O:** 77.8%, **N:** 80.0% ***MAO Inhibitor (PD only):** **O:** 68.1%, **N:** 71.0% ***Anticholinergic (PD only):** **O:** 8.3%, **N:** 8.7% ***Warfarin:** **O:** 14.5% & 1.4%, **N:**16.4% & 1.5% ***Statin:** **O:** 54.0% & 20.9%, **N:** 54.0% & 19.0%	**Oral:** 50.0 (C) & 51.4 (PD) **Nasal:** 50.8 (C) & 49.3 (PD)	**Oral:** 64.3 ± 7.0 (C) & 65.4 ± 5.5 (PD) **Nasal:** 64.4 ± 6.8 (C) & 65.3 ± 5.6 (PD)

C: controls. ***PD***: Parkinson's patients. ***O***: oral. ***N***: nasal. ***TIA***: Transient Ischemic Attack. ***IBS***: Irritable Bowel Syndrome. ***** denotes statistical significance in C vs. PD. Only the more relevant clinical variables are depicted in this table. For full details consult the supplementary material of the respective articles. All but two PD patients were under one or more antiparkinsonian medications, and two PD patients were treated by Deep Brain Stimulation (DBS).

3.1.1.1. Gut Study

Faecal samples were collected by the study subjects at home using specialised collection tubes pre-filled with Stool DNA Stabilizer (as part of the PSP® Spin Stool DNA Plus Kit package from STRATEC Molecular). Transfer to freezers at -80 °C occurred within three days of stool collection. Bulk DNA was extracted from the stool samples (PSP® Spin Stool DNA Plus Kit from STRATEC Molecular, as per the manufacturer's instructions), and PCR was performed in two separate rounds of amplification. During the first round, we used "universal" bacterial primers to amplify the V1-V3 regions of the 16S rRNA gene (Edwards *et al.* 1989, Lane 1991), with pA and pD' primers containing 18-mer overhangs in their 5' ends. Between 3.6 and 60 ng of template DNA were used in this first PCR round for each of three technical replicate reactions per sample, with a master mix containing Phusion polymerase (Finnzymes / Thermo Scientific), HF buffer, and 2.5% DMSO. The replicate products were then pooled and purified with Agencourt® AMPure® XP magnetic beads (Beckman Coulter). A second PCR round was then performed in which primers containing complementary sequences to the 18-mer overhangs are added, along with the Roche 454 adapter sequences added to the 5' ends and followed by an index in the pA primer. The cycling conditions for both PCR reactions are the following: initial denaturation at 98 °C for 30 s, 15 cycles at 98 °C for 10 s, 65 °C for 30 s, and 72 °C for 10s, followed by a final extension at 72 °C for 5 min. The final products were then purified again with Agencourt® AMPure® XP magnetic beads. An equimolar DNA solution was then prepared for sequencing using the 454-GS FLX Titanium chemistry (Roche Diagnostics).

3.1.1.2. Oral and Nasal Study

The wet lab protocol was similar to the one in section 3.1.1.1 but with the following changes: oral and nasal samples were collected using sterile cotton swabs at the clinic, and these were immediately placed into small containers in ice, followed by transfer to freezers at -80 °C within 20 min. Bulk DNA was extracted from both sample types with FastDNA™ Spin Kit for Soil (MP Biomedicals). The same universal bacterial primer sequences were used as in 3.1.1.1. On the 5' ends of the primers we added the full forward and reverse Illumina sequencing primers, for pA and pD' respectively, that served as overhangs for the second PCR round. Between 10.7 and 158.2 ng of template DNA were used in the first PCR round for each of three technical replicate reactions for the oral samples, and between 8.5 and 262.6 ng for the nasal samples. The master mix contained Phusion Hot-Start II polymerase (Finnzymes / Thermo Scientific). Purification of the pooled technical replicates from the first PCR round was performed by enzymatic digestion with Exonuclease I and FastAP (Thermosensitive Alkaline Phosphatase, Thermo Scientific). The second PCR round used primers with Illumina adapters at the 5' ends followed by partial Illumina sequencing primer sequences that will attach to the overhangs from the previous PCR round. Index sequences were built into the reverse primers. The cycling conditions for both PCR rounds were similar to the ones described in section 3.1.1.1 except for the number of cycles in each of the rounds: the first used 15 cycles and the second one 18. The final equimolar DNA pool was then prepared for sequencing on an Illumina MiSeq using the v2 600 cycle kit paired-end (325 bp + 285 bp).

3.1.2. Primary Sclerosing Cholangitis

Table 2 presents general clinical data for the study populations in Study III. Only the main clinical variables are presented here. For more information consult Article III and its supplementary material.

Table 2. Selected clinical data for Study III.

Selected Clinical Variables	Controls (n = 46)	Early Disease Stage (n = 37)	Advanced Disease Stage (n = 32)	Dysplasia / Carcinoma (n = 11)
Female, n/total (%)	27/46 (59)	25/37 (68)	17/32 (50)	9/11 (82)
Age, years, mean (SD)	43 (15)	38 (14)	41 (13)	40 (15)
***IBD present, n/total (%)**	9/46 (20)	25/37 (68)	20/32 (59)	8/11 (73)
***ERC Score, 0-16 (SD)**	0 (0.0)	2.8 (1.1)	7.9 (1.8)	10.1 (2.3)
***No of ERCs, n (%):**	-	-	-	-
1-2	[a]46/46 (100)	34/37 (92)	14/32 (41)	4/11 (36)
3-4	0/46 (0)	3/37 (8)	8/32 (24)	4/11 (36)
≥ 5	0/46 (0)	0/37 (0)	12/32 (35)	3/11 (27)

IBD: Inflammatory Bowel Disease. ***ERC***: Endoscopic Retrograde Cholangiography. * denotes statistical significance between groups. [a] denotes only one ERC examination. Only the more relevant clinical variables are depicted in this table. For full details consult Table 1 in Article III.

Bile fluid was obtained during ERC examinations using aspiration and divided into 1 ml plastic tubes that were immediately immersed in liquid nitrogen at -196 °C and then placed in freezers at -20 °C. Later they were moved to -80 °C freezers until further processing. Bulk DNA was extracted using Invisorb® Spin Blood Mini Kit (Stratec Molecular). The PCR procedure was similar to the one in section 3.1.1.2, but with two technical replicates per sample. Between 2.8 and 352 ng of template DNA were used during the first PCR round per technical replicate. The final equimolar pool was sequenced using the Illumina MiSeq v3 600 cycle kit paired-end (325 bp + 285 bp).

3.2. Data Analysis: Overview of the Methods

After the biological samples are sequenced and the raw sequence data is obtained, the latter needs to be processed to produce biologically meaningful information that can be analysed by the end user. This essentially means converting the raw sequence data into lists of organisms present in their respective communities, including their relative abundances. Different software exists to perform these bioinformatic analyses. In microbial ecology and human microbiome studies the two most used software packages are *QIIME* (Caporaso *et al.* 2010) and *mothur* (Schloss *et al.* 2009). These are software packages that allow an entire sequence processing pipeline to be performed without resorting to handling separate programs to perform specific functions. Instead, these generalist packages contain implementations of open source software under the same user interface. *mothur* was used to handle all the sequence data in the three studies presented in this thesis (I, II, III). The main sequence processing steps in *mothur* consist of, in order:

1) Reduction of sequencing errors (removal of sequences with ambiguous base calls, removal of homopolymers above a defined maximum length, minimum average quality score per sequence, detecting and correcting sequencing errors in the form of base substitutions, etc). Reducing sequencing errors is important, since sequencing errors are difficult to distinguish from true biological variation, thus artificially inflating microbial diversity (Quince *et al.* 2011, Schloss *et al.* 2011, Bragg *et al.* 2012, Kozich *et al.* 2013).

2) Alignment of the dataset to an external database (we used SILVA, the default; Quast *et al.* 2013, Schloss 2009), and removal of sequences not overlapping with the target regions (V1-V3 regions of the 16S rRNA gene in our three studies).

3) A final step in denoising the dataset can be performed at this point, by pre-clustering the sequences, followed by a pre-specified similarity threshold for sequence merging. We set the parameter to 4 instead of 2 nucleotides difference between a given sequence and a second, more abundant sequence, since our V1-V3 region is longer than the one described in the online protocol (circa 500bp).

3) Identification and removal of chimeric sequences produced by PCR amplification, either during DNA library creation or potentially in the sequencing platform. We used the most abundant sequences as reference, which is the default approach, as opposed to using an external database e.g. SILVA (Meyerhans *et al.* 1990, Smyth *et al.* 2010, Edgar *et al.* 2011).

4) Removal of undesirable sequences derived from chloroplasts, mitochondria, *Archaea*, and eukaryotes (18S rRNA) that get accidentally picked by the bacterial primers due to homology. The identification and filtering of these sequences is done by aligning the reads to a reference database e.g. from the Ribosomal Database Project (RDP; Cole *et al.* 2009).

5) An optional step at this stage is the removal of singleton sequences. It helps with processing power and also reduces the amount of false biological reads produced by sequencing errors e.g. low quality sequences at the end of the reads. If processing power isn't an issue, removal of singletons can be performed prior to statistical analysis. In Study I we did not remove the singleton sequences at this stage, but instead prior to statistical analysis. Also, we used the *phylotype approach* in this study (see below), which produces less singletons. In the remaining two studies (II & III) we removed the singletons at this stage.

6) Clustering and taxonomic classification of the remaining sequences. For Study I we used the *phylotype approach*, and for the remaining two studies (II & III) we used the *OTU-based approach*. The phylotype approach relies on both binning and classifying the sequences according to sequence similarity to an external, curated database. We used the RDP database. The OTU-based approach relies on binning the sequences first according to similarity to each other using the dataset as self-reference and then using an external database for the taxonomic classification. We also used RDP in this case (Schloss & Westcott 2011, Rideout *et al.* 2014, Westcott & Schloss 2015).

In Study I, using the Roche 454 platform, we followed the instructions provided in the protocol by Schloss, publicly available online (Schloss *et al.* 2011; Schloss, access date July 2013). In studies II & III we used the Illumina MiSeq protocol, also freely available online (Kozich *et al.*

2013; Schloss, access dates October 2014 and June 2014, respectively). For a detailed treatment of sequence processing and related matters, see Meyerhans *et al.* 1990, Schloss 2009, Schloss 2010, Smyth *et al.* 2010, Edgar *et al.* 2011, Quince *et al.* 2011, Schloss *et al.* 2011, Schloss & Westcott 2011, Bragg *et al.* 2012, Schloss 2013, Koskinen *et al.* 2015, Westcott & Schloss 2015.

For studies II & III we sequenced DNA extraction kit blanks and PCR blanks as contamination controls. This resulted in the removal of all sequences from the genera *Halomonas*, *Ralstonia*, and *Shewanella*, based on the presence and amount of reads on both control types and on published data on contaminants, as well as on personal experience with previous studies at our lab (Salter *et al.* 2014, Aho *et al.* 2015).

Statistical data analysis was performed with the *R* statistical programming language (R Core Team 2015) except otherwise noted. The package *phyloseq* (McMurdie & Holmes 2013) was used to manipulate the microbiome data in *R*'s environment. All *p*-values were double-tailed, with statistical significance accepted at an alpha of 0.05. In Article I we used SPSS Statistics (IBM Corporation) for Generalised Linear Models (GLMs) and Receiver Operating Characteristic curves (ROC curves).

In the context of microbiome studies, alpha diversity is an estimation in the form of an index that takes into account the number of different species or OTUs (*richness*) and how well represented they are relative to each other (*evenness*) in any given biological sample. We used mostly the Shannon and the inverse Simpson diversity indices in our work (Haegeman *et al.* 2013). Study I was performed with rarefied data after removal of singletons. Studies II & III were performed using non-rarefied data after removal of singletons and decontamination. Although greater sequencing depth in any given sample has the potential to increase the number of OTUs observed (Weiss *et al.* 2017), these differences will balance out in practice for purposes of alpha diversity between-group comparisons (personal observation). Removal of singletons for calculation of diversity indices is usually frowned upon by some authors, but there are fundamental differences between the use of these indices in the traditional ecological literature and in DNA sequencing-based microbial ecology. These indices were originally developed with macro-organisms in mind (Whittaker 1960), where taxa can be identified with less ambiguity than phylogenetic marker genes. Also, given that many of these singleton sequences are the product of sequencing errors that can create artificial OTUs, it is desirable to remove them. Absolute diversity indices in microbial marker-gene surveys can only be effectively compared within-study, and thus a "true" index value valid for inter-study comparisons is essentially meaningless. Differences in alpha diversity were tested with Kruskal-Wallis rank sum tests and/or Pairwise Wilcoxon rank sum tests, with False Discovery Rate (FDR) correction for multiple comparisons. These tests are non-parametric and adequate for data whose distribution deviates from normality and, more importantly, for data that is heteroskedastic.

Beta diversity measures the composition of microbial communities (*what* species or OTUs are present and in what numbers) in the groups under contrast (*sites*, in the traditional ecological literature; Anderson *et al.* 2011). We used the Bray-Curtis dissimilarity index for these calculations, which produces a dissimilarity matrix based on pairwise sample differences, visualised the matrix data with Non-Metric Multidimensional Scaling (NMDS), and tested for differences between groups with the *adonis* function of the package *vegan* (Oksanen *et al.*, access date 2015). For Article I, we used UNIFRAC in *mothur* instead of *adonis* (Lozupone & Knight 2005, Lozupone et al. 2011). *adonis* is an implementation of PERMANOVA (Permutational MANOVA using Distance Matrixes), formerly known as Non-Parametric MANOVA (Anderson 2001, McArdle & Anderson 2001). The data was rarefied before the estimations, since these will be sensitive to the presence or absence of particular taxa across compared groups. We informally

tested the latter and in some cases non-rarefied data leads to artefacts such as samples with more reads clustering together, very probably due to containing more (usually rare) taxa in common. Thus, we use rarefaction as a precaution, and in fact this has been shown to be a common problem with compositional data in beta diversity analysis (Weiss *et al.* 2017).

NMDS plots and *adonis* were also used to investigate potential batch effects resulting from DNA extraction batches, kit lot batches, PCR batches, and multiple sequencing runs. When the existence of potential effects was suggested by the analysis, we controlled for those effects during modelling (McNamee 2005, Leek *et al.* 2010). For a discussion of common problems with contamination and batch effects in 16S rRNA gene survey studies see e.g. Salter *et al.* (2014).

Collinearity and multicollinearity (when two or more explanatory variables in a model are substantially correlated with each other, respectively) were assessed with pair-plots of Pearson Correlation coefficients and with Variance Inflation Factors (VIF), for comparisons between continuous variables. For collinearity between continuous and categorical variables with more than two levels we used boxplots and Generalised VIFs. In case (multi)collinearity was detected, adequate variable selection was undertaken before running the GLMs (Ieno & Zuur 2015, Fox 2016).

Logistic regression was used for prediction of PD status based on bacterial abundances and other explanatory variables in Article I. ROC curves were used to assess the discriminative performance of the various classifier models in terms of sensitivity (true positive rate) and specificity (true negative rate; Loong 2003).

Differential abundance analysis was performed with *Metastats* in Article I (White *et al.* 2009), using *mothur*. Given that *Metastats* does not allow adjustment for potential confounders and consists, in general terms, in directly comparing pre-defined groups/variables of interest with the non-parametric *t*-test (Storey & Tibshirani 2003) and Fisher's exact test for rare taxa, both with FDR, we had to further evaluate the statistically significant hits preliminarily obtained from *Metastats* with GLMs. Taxa that were statistically significant before correction for multiple comparisons with the FDR method were also further investigated with GLMs.

In studies II & III, differential abundance analysis was performed with Negative Binomial GLMs using the *DESeq2* package for *R* (Love *et al.* 2014). Instead of rarefying the raw count data, *DESeq2* normalises the data for sequencing depth, without the losses common to rarefaction (Hanski *et al.* 2012, McMurdie & Holmes 2014), using the *median-of-ratios* of observed counts' method (Anders & Huber 2010), from which it estimates *size factors*. These size factors are included as offsets directly in the models, and therefore there is no data normalisation prior to running the models. *DESeq2* uses Logarithmic Fold Change (LFC) estimation for effect sizes: the null hypothesis for differential abundance testing is that the LFC (not the mean abundance of counts) between two statistical samples is zero. These are estimated during a first round of model fitting using Maximum-Likelihood Estimation (MLE) of the log fold changes, using only individual taxon data. Because data with high within-group variances can lead to heteroskedasticity, to statistically significant LFC estimates due to chance effects, or to biologically meaningless large fold change estimates, *DESeq2* shrinks these raw LFCs towards zero. The greater the within-group variance, the more shrinkage is applied to a taxon's LFC. The software does this by calculating a prior distribution, based on the distribution of all observed LFCs across the dataset and to which a normal distribution centred on zero is fit. This prior distribution is then used as a Bayesian prior (*Empirical Bayes Shrinkage*) for estimating the final, adjusted LFCs when refitting a second round of GLMs per individual taxon (Love *et al.* 2014).

Multiple comparisons' correction was done using the Benjamin-Hochberg method (an FDR procedure) to avoid Type-I errors. We also filtered the data before running the GLMs as

an added precaution against biologically meaningless large fold changes or chance effects driven by low count taxa or by groups with low sample numbers (Giesecke 2002, Freedman 2009). This filtering was performed by removing from the dataset, prior to running the models, all taxa that weren't represented by at least 2 sequences per biological sample in at least 11 biological samples across the dataset. This is a conservative approach, since *DESeq2* performs automatic filtering. This procedure may lead to the unintended purge of potential keystone species from the dataset, but in the present framework it is a necessary compromise due to the need to minimize Type I errors.

Oligotyping was used to investigate the *Staphylococcus* OTU detected in the nasal data of Article II. The analysis was performed following the pipeline provided by the authors' in Eren *et al.* (2013). Oligotyping is used to help differentiate between taxa obtained from 16S rRNA gene amplicons, since these may be unresolved at the desired taxonomic level due to limitations inherent in the sequence classification methods available (e.g sequence similarity thresholds for clustering sequences into OTUs and/or incomplete reference databases). Unlike reference-based or clustering-based methods that compare all available sequence positions at the same time, oligotyping uses only the variable sites, after being identified by entropy analysis, that optimize taxonomic granularity by focusing on information-rich nucleotide variation. The method does not depend on similarity of conserved regions of entire sequences, considering these redundant for finer taxonomic discrimination. Thus, the method will produce differentiated operational units, oligotypes, that are based on the minimal number of nucleotide positions in the taxon/operational unit of interest that will explain the maximum amount of biological diversity within that taxon/operational unit (e.g. an OTU sequence cluster or all sequences classified as a particular genus). These oligotypes can then be subjected to differential abundance analysis. Their sequences can also be matched to a database and cladograms can be built to elucidate phylogenetic relationships.

The statistically significant taxa resulting from the differential abundance models were further investigated using boxplots and scatterplots to look for signs of presence of chance effects that could result in Type-I errors. This was done by plotting the variable of interest and the relative abundance (or the normalised count data) of the taxon in question and investigating the patterns and spread of the data, e.g. for the presence of putative outliers, influential points with high leverage, very low mean abundances (e.g. of 1) regardless of effect size, inconsistent patterns of increasing and decreasing abundance across multiple groups that are difficult to explain as biologically meaningful, etc. These features are usually found in combination. When statistical significance was suspected of being just an apparent effect derived from statistical noise, we took a conservative approach and removed the corresponding taxa from further consideration. Tables with all the obtained statistically significant results were nevertheless provided, either in the Articles' main text or in their respective supplementary materials.

4. RESULTS

4.1. Parkinson's Disease

With data from a recent, unpublished gut dataset using Illumina MiSeq sequencing and the same stool samples and subjects from Study I, we generated an NMDS plot to contrast the compositional data from gut, oral, and nasal samples (Figure 3). As can be seen, there is a clear compositional demarcation based on sampling location.

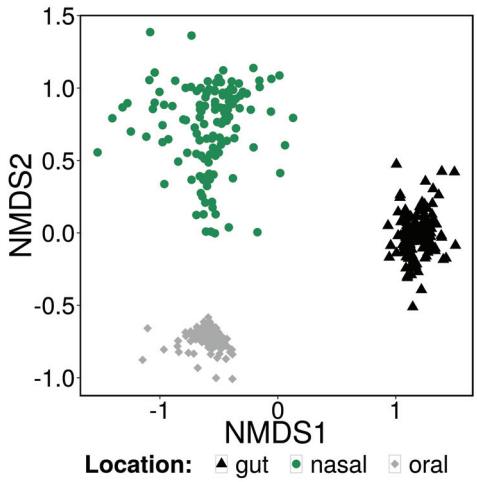


Figure 3. NMDS plot of PD data by sampling location. Ordination based on Bray-Curtis dissimilarity with genus-level data. Each point represents one subject.

4.1.1. Gut Study

The original dataset consisted of 2 549 217 raw reads. The final dataset used for statistical analysis after processing consisted of 1 131 504 reads, with a mean of 7 645 reads per subject. The overall taxonomic composition of the dataset reflects the results of other studies on gut microbiota (e.g. Qin *et al.* 2010, Hasegawa *et al.* 2015, Keshavarzian *et al.* 2015). We detected 18 phyla, 29 classes, 60 orders, 125 families, and 360 genera of bacteria, with the most common phyla being represented by *Firmicutes* and *Bacteroidetes*, which are the typically dominant phyla in the gut (Figure 4).

No differences in alpha diversity were found between the PD and control groups, suggesting that OTU richness and evenness are similar among the two populations (mean inverse Simpson index for controls = 20.6 *vs.* PD subjects = 21.8, $p = 0.38$; mean Shannon index for controls = 3.7 *vs.* PD subjects = 3.8, $p = 0.51$, with Kruskal-Wallis rank sum test). On the other hand, beta diversity analysis suggested the existence of compositional differences at family-level between the bacterial communities of the two groups (Yue & Clayton theta, Morisita-Horn, and Bray-Curtis indices computed with family-level data; unweighted UniFrac $p \leq 0.02$ and weighted UniFrac $p < 0.001$ for all three indices).

Metastats identified six families that were potentially differentially abundant, namely *Prevotellaceae*, *Lactobacillaceae*, *Verrucomicrobiaceae*, *Bradyrhizobiaceae*, *Clostridiales Incertae Sedis IV*, and *Ruminococcaceae* (Table 3). All were estimated as increasing in abundance in the PD group except for *Prevotellaceae*. The latter was the fourth most abundant family in the dataset

and showed the largest abundance differential: a decrease of 77.6% in PD relative to the control group (Figure 4, Table 3), with subjects with high abundances of *Prevotellaceae* not being present in the PD group (Figure 1 in Article I).

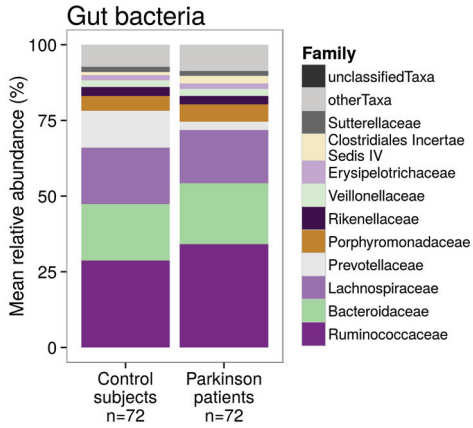


Figure 4. Relative abundances of the 10 most abundant bacterial families in the dataset.
The 10 and 5 most common families accounted for 91.6% and 81.3% of all reads, respectively.

Table 3. Metastats results. Family differential abundances, PD vs. Controls.

Family	Patients	Controls	*p*-value	Q-value
Prevotellaceae	2.70 ± 0.66	12.06 ± 2.28	0.001	0.031
Lactobacillaceae	0.44 ± 0.23	0.02 ± 0.01	0.004	0.063
Verrucomicrobiaceae	0.06 ± 0.02	0.02 ± 0.00	0.014	0.146
Bradyrhizobiaceae	0.16 ± 0.06	0.03 ± 0.01	0.021	0.151
Clostridiales Incertae Sedis IV	2.49 ± 0.65	1.01 ± 0.22	0.025	0.151
Ruminococcaceae	33.63 ± 1.66	28.54 ± 1.58	0.029	0.151

PD and control groups with an uncorrected p-value less than 0.05 (mean % ± standard error).
Q-value: correction for multiple comparisons (FDR).

Given that *Metastats* does not permit adjustment for potential confounders based on the clinical data, we used GLMs with the six families above as response variables (Tables 1 and 2 in Article I). *Prevotellaceae* were shown to vary in abundance only in response to disease status, with no other model variable showing any detectable effect. All other families, except for *Ruminococcaceae*, were associated with disease status as well as with other clinical factors. *Ruminococcaceae* were found to be strongly associated only with *Prevotellaceae* abundance levels and not with disease status or any other clinical variable.

A ROC curve analysis based only on *Prevotellaceae* abundances had a sensitivity of 86.1% and a specificity of 38.9% for identification of PD patients (Figure 2 in Article I). Using logistic regression with disease status as response variable and *Prevotellaceae*, *Lactobacillaceae*, *Bradyrhizobiaceae*, and *Clostridiales Incertae Sedis IV* abundances, as well as the Wexner total score (a clinical measure of constipation) as explanatory variables, enabled identification of PD cases with a sensitivity of 66.7% and a specificity of 90.3%.

We made a PD subjects-only GLM analysis based on the UPDRS-III total score (Unified Parkinson's Disease Rating Scale; Part III refers to the clinician-scored motor evaluation), including subjects under normal medication and DBS (Deep Brain Stimulation). The response variables were the five families previously found to be statistically significantly associated with disease status. Although *Prevotellaceae* were statistically significantly, and positively, correlated with increasing UPDRS-III severity, the statistical significance was lost when subjects with motor fluctuations and under DBS were removed from the analysis.

Metastats identified *Enterobacteriaceae* as differentially abundant between the TD (Tremor Dominant) and PIGD (Postural Instability and Gait Difficulty) groups within the PD subjects, showing an increase in abundance in the PIGD group (Table 4, eFigure 2 in Supplement for Article I). This family was further investigated with two GLMs to evaluate possible correlations with the Tremor subscore, PIGD subscore, Akinetic-rigid subscore, and the NMSS total score (Non-Motor Symptoms Scale; Table 3 in Article I). *Enterobacteriaceae* abundance was found to be positively correlated with the PIGD and Akinetic-rigid subscores, and was very close to statistical significance for a negative correlation with the Tremor subscore. Removal of subjects with motor fluctuations or under DBS from the analysis results in only the PIGD correlation remaining significant.

Table 4. Metastats results. Family differential abundances, TD vs. PIGD.

Family	TD	PIGD	*p*-value	Q-value
Enterobacteriaceae	0.28 ± 0.13	2.31 ± 0.81	0.004	0.018
Erysipelotrichaceae	1.17 ± 0.30	2.19 ± 0.32	0.024	0.103

TD and PIGD groups with an uncorrected p-values less than 0.05 (mean % ± standard error).
Q-value: correction for multiple comparisons (FDR).

4.1.2. Oral and Nasal Study

Alpha diversity was higher in the oral than the nasal dataset. Beta diversity analysis showed that the oral and nasal community compositions were very distinct (Figure 3). Also, the oral and nasal samples from any given individual (irrespective of PD status) weren't more similar to each other than those of unrelated subjects (Wilcoxon rank sum test, $p = 0.88$ for Bray-Curtis dissimilarity between oral and nasal samples *vs.* unrelated pairs of samples; Figure S2 in the supplementary data for Article II), with the samples being more similar according to sampling location than to particular individuals (Figure 1 in Article II).

4.1.2.1. Oral

The original oral dataset consisted of 21 645 150 raw reads. The final dataset used for statistical analysis after processing consisted of 1 987 257 reads, with a mean of 13 427 reads per subject. We detected 15 phyla, 23 orders, 41 classes, 76 families, 129 genera, and 7996 OTUs. The taxonomic composition of the oral communities inferred from the data was consistent with previous findings (e.g. Bik *et al.* 2010, Palmer 2014), with *Streptococcus* seemingly dominating oral communities (Figure 2A in Article II).

No difference was detected in alpha diversity between PD subjects and controls (mean Shannon index for controls = 2.69 *vs.* PD subjects = 2.75, mean inverse Simpson for controls = 7.70 *vs.* PD subjects = 9.16; Kruskal-Wallis rank sum test, $p \geq 0.86$ for both indices). Beta diversity analysis suggested the presence of differences in community composition between controls and PD (*adonis*, $p = 0.01$), and between genders regardless of PD status (*adonis*, $p = 0.05$).

The most common genera in the oral dataset were *Streptococcus*, *Haemophilus*, *Neisseria*, *Veillonella*, and *Prevotella* (Figure 2A in Article II). PD status was the GLM variable associated with the greater number of detected taxon differential abundances, namely 11 families, 10 genera, and 25 OTUs (Table 5). Many of these taxa contain members that are known to be opportunistic oral pathogens. An increase in potential oral pathogen abundances was also detected in males relative to females, regardless of PD status. Minor results for other variables can be seen in Table 2 of Article II, as well as in the supplementary tables for Article II.

Table 5. Differentially abundant (DESeq2: $p < 0.05$) taxa for oral data, Controls -> PD.

Taxon	Mean Abundance	Log 2 Fold Change	SE of Log 2 Fold Change	*p*-value	Adjusted *p*-value
OTU					
Otu000008 (Haemophilus)	419.00	-1.7031	0.3677	3.63E-06	1.94E-04
Otu000011 (Neisseria)	453.14	-2.6123	0.5175	4.46E-07	1.67E-04
Otu000009 (Gemella)	324.27	-1.0299	0.2745	1.76E-04	5.50E-03
Otu000017 (Rothia)	187.36	-1.282	0.4064	1.61E-03	2.51E-02
Otu000026 (Veillonella)	77.23	1.5349	0.3785	5.01E-05	2.09E-03
Otu000028 (Prevotella)	73.46	1.1903	0.3528	7.42E-04	1.48E-02
Otu000073 (Leptotrichia)	38.50	-1.801	0.5291	6.64E-04	1.47E-02
Otu000074 (Prevotella)	24.77	1.6926	0.5261	1.29E-03	2.11E-02
Otu000123 (unclassified Flavobacteriaceae)	24.14	-1.5284	0.3142	1.15E-06	1.94E-04
Otu000127 (Corynebacterium)	21.66	-1.65	0.4696	4.42E-04	1.11E-02
Otu000118 (Kingella)	19.83	-1.5231	0.452	7.52E-04	1.48E-02
Otu000107 (Prevotella)	18.72	1.6819	0.3605	3.07E-06	1.94E-04
Otu000212 (Kingella)	9.50	-3.4138	0.7324	3.15E-06	1.94E-04
Otu000210 (Neisseria)	8.95	-2.8492	0.6129	3.34E-06	1.94E-04
Otu000231 (Corynebacterium)	8.11	-2.2211	0.5861	1.51E-04	5.15E-03
Otu000265 (Neisseria)	8.08	-2.3153	0.5992	1.12E-04	4.19E-03
Otu000352 (Neisseria)	4.77	-2.2078	0.6187	3.59E-04	9.62E-03
Otu000361 (Neisseria)	4.63	-2.0676	0.6323	1.08E-03	1.97E-02
Otu000355 (Neisseria)	4.52	-2.4536	0.595	3.73E-05	1.75E-03
Otu000367 (Actinomyces)	3.96	-2.0336	0.624	1.12E-03	1.97E-02
Otu000509 (Capnocytophaga)	2.05	-3.5443	0.9682	2.51E-04	7.25E-03
Otu000392 (Solobacterium)	1.86	1.3991	0.4079	6.04E-04	1.42E-02
Otu000448 (Kingella)	1.58	-4.9599	1.0673	3.37E-06	1.94E-04
Otu000418 (Prevotella)	1.55	2.3684	0.7288	1.15E-03	1.97E-02
Otu000806 (Veillonella)	0.28	2.0303	0.6551	1.94E-03	2.91E-02

Table 5 cont.

Taxon	Mean Abundance	Log 2 Fold Change	SE of Log 2 Fold Change	*p*-value	Adjusted *p*-value
Genus					
Haemophilus	1356.82	-1.021	0.3477	3.32E-03	2.07E-02
Gemella	673.01	-0.8461	0.3261	9.46E-03	4.73E-02
Neisseria	666.52	-1.4365	0.3233	8.87E-06	2.22E-04
Veillonella	410.55	0.7864	0.2301	6.31E-04	5.26E-03
Prevotella	283.03	0.9594	0.2408	6.76E-05	8.45E-04
Rothia	126.94	-1.1407	0.3106	2.40E-04	2.40E-03
Granulicatella	88.27	-0.715	0.2606	6.09E-03	3.38E-02
Kingella	28.79	-1.3791	0.3022	5.04E-06	2.22E-04
Moryella	2.34	0.9946	0.3175	1.74E-03	1.24E-02
Solobacterium	1.59	1.3087	0.3146	3.18E-05	5.30E-04
Family					
Pasteurellaceae	2718.29	-0.8065	0.3292	1.43E-02	4.97E-02
Neisseriaceae	710.01	-1.3765	0.2851	1.38E-06	4.83E-05
Veillonellaceae	458.86	0.7763	0.2146	2.97E-04	2.60E-03
Prevotellaceae	349.18	1.0504	0.2351	7.86E-06	1.38E-04
Leptotrichiaceae	287.61	-0.5664	0.2099	6.96E-03	3.48E-02
Micrococcaceae	127.91	-1.1177	0.2942	1.45E-04	1.70E-03
Carnobacteriaceae	88.80	-0.789	0.2471	1.41E-03	8.20E-03
Corynebacteriaceae	77.91	-0.7095	0.2695	8.47E-03	3.71E-02
Lactobacillaceae	10.00	1.1188	0.4628	1.56E-02	4.97E-02
Coriobacteriaceae	5.84	0.7895	0.3128	1.16E-02	4.51E-02
Erysipelotrichaceae	1.84	0.9778	0.283	5.50E-04	3.85E-03

Model variables: taxon ~ extraction date + gender + age at sampling + history of ENT surgery + alcohol units per week + SDQ total + SCS-PD total + control vs PD.

4.1.2.2. Nasal

The original nasal dataset consisted of 8 638 162 raw reads, and the final dataset for statistical analysis consisted of 2 610 933 reads, with a mean of 19 198 reads per subject. We detected 28 phyla, 49 classes, 96 orders, 177 families, 553 genera, and 2212 OTUs. The taxonomic composition of the nasal communities inferred from the data was consistent with that reported in previous findings (Bassis *et al.* 2014, Biswas *et al.* 2015), with the dominant genus being *Corynebacterium* (Figure 2C in Article II).

No differences were detected in alpha diversity between controls and PD subjects (mean Shannon index for controls = 1.75 *vs.* PD subjects = 1.80, mean inverse Simpson for controls = 4.86 *vs.* PD subjects = 4.69; Kruskal-Wallis rank sum test, $p \geq 0.65$ for both indices). Notice that the mean values for the indices are not directly comparable between the oral and nasal datasets, since the raw sequence data was processed separately in *mothur*. Beta diversity analysis with *adonis* suggested that no compositional differences are present between controls and PD subjects.

The most common genera in the nasal dataset were *Corynebacterium*, *Propionibacterium*, *Moraxella*, *Staphylococcus*, and *Burkholderia* (Figure 2B in Article II). Only one taxon per each of the three taxonomic levels investigated were differentially abundant for PD status: a *Staphylococcus* OTU increased in PD, the genus *Marmoricola*, decreased in PD, and the family Flavobacteriaceae, also decreased in PD (Table 6, and Figure S5 in supplementary data for Article II). Minor results for other variables can be seen in Table 2 of Article II, as well as in the supplementary tables for Article II. The *Staphylococcus* genus was further investigated using oligotyping, which produced three oligotypes, one resembling *S. epidermis*, another *S. aureus*, while the last one was an unresolved *Staphylococcus* (Figure S6 in Supplement for Article II). The first two composed 95% of all *Staphylococcus* reads. All were similar in relative abundance across PD and controls, producing no statistically significant results.

Table 6. Differentially abundant (DESeq2: $p < 0.05$) taxa for nasal data, Controls -> PD.

Taxon	Mean Abundance	Log 2 Fold Change	SE of Log 2 Fold Change	p-value	Adjusted p-value
Otu00006 (Staphylococcus)	1512.91	3.9094	0.7019	2.55E-08	4.64E-06
Genus Marmoricola	4.56	-5.4099	1.4188	1.37E-04	1.78E-02
Family Flavobacteriaceae	31.50	-2.795	0.7322	1.35E-04	1.08E-02

Model variables: taxon ~ DNA extraction batch + gender + age at sample collection + history of non-major ENT surgery + sniffin' sticks score + control vs PD.

4.2. Primary Sclerosing Cholangitis Study

The study subjects were classified as follows for the purpose of investigating the research questions: controls (ERC severity score of 0, first ERC examination), patients at an early disease stage (ERC severity score < 6), patients at an advanced disease stage (ERC severity score ≥ 6), and patients with biliary dysplasia/cholangiocarcinoma (Table 2).

The original bile dataset consisted of 22 795 599 raw reads, and the final dataset for statistical analysis consisted of 3 740 318 reads, with a mean of 29 686 reads per subject. We detected 20 phyla, 34 classes, 61 orders, 124 families, 309 genera, and 2125 OTUs. The most common phyla were *Bacteroidetes*, *Firmicutes*, and *Proteobacteria* (Figure 1a in Article III), and the most common genera in the dataset were *Prevotella*, *Streptococcus*, *Veillonella*, *Fusobacterium*, and *Haemophilus* (Figure 1c in Article III). Overall, these putative bile microbial communities show some similarity to gut communities (Hiramatsu *et al.* 2000, Folseraas *et al.* 2012, Shen *et al.* 2015, Ye *et al.* 2016).

Alpha diversity analysis with the Kruskal-Wallis rank sum test and Shannon index suggests that differences exist among the four subject groups ($p = 0.036$). To discriminate among the groups, we ran Pairwise Wilcoxon rank sum tests. These suggest that the differences are associated mainly with controls *vs.* the dysplasia/carcinoma group ($p = 0.055$), and the early disease stage and the dysplasia/carcinoma group ($p = 0.055$; Figure 2a in Article III). All other comparisons had $p = \geq 0.164$. The overall pattern suggests a progressive decrease in alpha diversity, with controls and early disease stage patients being virtually identical. The latter observation is substantiated

by a Pairwise Wilcoxon rank sum test using controls against early disease stage patients that had undergone only one ERC examination, to avoid potential ERC effects (p = 0.64; Figure 2b in Article III). Beta diversity analysis suggested the presence of differences in community composition among the four study groups when the latter are not discriminated based on the number of ERC examinations (*adonis*, p = 0.05). When the groups are stratified by number of ERC examinations (one *vs.* multiple) no differences were detected (*adonis*, p = 0.34), suggesting an effect of ERC examinations on community composition. Finally, both the ERC severity score and the number of ERC examinations seem to have an impact on community composition (p = 0.06 and p = 0.02, respectively).

Differential abundance analysis resulted in a long list of taxa with statistical significance for each of the groups and variables compared (Table S1 in Supplement to Article III). Comparing controls to early disease patients at their first ERC examination resulted in four OTUs: an unclassified *Enterobacteriaceae* (Otu0008), an unclassified *Neisseriaceae* (Otu00213), *Neisseria* (Otu00045), and *Campylobacter* (Otu00089), as well as three families, *Pasteurelaceae*, *Staphylococcaceae*, and *Xanthomonadaceae* (Table 7). With controls and the early disease group as a whole (adjusting for the number of ERC examinations) we obtained two OTUs, an unclassified *Clostridiales*, Otu00188, and the same unclassified *Neisseriaceae* as before, Otu00213, and one family, *Staphylococcaceae*. However, visual inspection, very low abundances, and LFC considerations suggest that most of these results are Type-I errors (see also Figures S1 and S2 in Supplement to Article III).

The remaining models and comparisons identified 24 taxa for early disease *vs* advanced disease stages, and 36 taxa for advanced disease *vs* dysplasia/carcinoma stages (Table S1 in Annex X to Article III).

Most of these statistically significant taxa also showed a combination of features that renders them unconvincing as real biological effects. Therefore, we decided to concentrate our attention only on those results that looked robust after assessment, although it is plausible that later studies may validate some of those results. The only clear, robust results refer to the streptococcal group and to a *Prevotella* OTU (Table 7, Figure 3a in Article III). From the early disease to the advanced disease group we saw an increase in the abundance of one *Streptococcus* OTU (Otu00020) and of the genus itself. From the advanced disease to the dysplasia/carcinoma groups we saw a rise in abundance of two different *Streptococcus* OTUs (Otu00035 and Otu00061). The *Prevotella* OTU (Otu00128) had low mean abundance but a clear pattern, virtually disappearing in the dysplasia/carcinoma group.

When the disease severity was measured with the ERC severity score as opposed to a categorical variable stratified by disease stage, we detected two *Streptococcus* OTUs (Otu00020 and Otu00061), each of them in common with one of the previous comparisons, and both positively correlated with increasing disease severity (Table 7, Figure 3b in Article III).

The number of ERC examinations was included in the models and also investigated, since previous publications reported a potential impact of this procedure on the microbiota, especially on the streptococcal group. The results showed that one *Streptococcus* OTU (also detected in the dysplasia/carcinoma comparison, Otu00035) as well as the genus itself are positively correlated with the number of ERC examinations undergone by the patients (Table 7).

Table 7. Differentially abundant (DESeq2: $p < 0.05$) taxa for bile data.

Family	Genus	OTU	Mean Abundance	Log 2 Fold Change	SE of Log 2 Fold Change	*p*-value	Adjusted *p*-value
Controls -> patients with early disease and no history of ERC examinations							
Enterobacteriaceae	unclassified	Otu00008	640	-4.65	1.27	2.42E-04	1.98E-02
Neisseriaceae	Neisseria	Otu00045	33	-3.18	0.94	7.14E-04	3.78E-02
Campylobacteraceae	Campylobacter	Otu00089	11	-2.79	0.77	2.80E-04	1.98E-02
Neisseriaceae	unclassified	Otu00213	1	-5.01	1.17	1.78E-05	3.78E-03
Pasteurellaceae	-	-	14058	1.95	0.64	2.12E-03	3.25E-02
Staphylococcaceae	-	-	179	3.97	0.76	1.73E-07	7.95E-06
Xanthomonadaceae	-	-	11	3.58	1.11	1.20E-03	2.76E-02
Controls -> patients with early disease							
unclassified Clostridiales	unclassified	Otu00188	1	-6.10	1.41	1.43E-05	1.52E-03
Neisseriaceae	unclassified	Otu00213	1	-4.91	1.10	7.64E-06	1.52E-03
Staphylococcaceae	-	-	179	4.18	0.76	4.09E-08	2.25E-06
Patients with early disease -> patients with advanced disease							
Streptococcaceae	Streptococcus	Otu00020	728	4.94	0.83	3.22E-09	3.93E-07
Streptococcaceae	Streptococcus	-	9241	1.44	0.42	6.70E-04	4.94E-03
Patients with advanced disease -> patients with dysplasia/carcinoma							
Streptococcaceae	Streptococcus	Otu00035	602	5.78	1.31	1.00E-05	7.10E-04
Streptococcaceae	Streptococcus	Otu00061	23	4.02	1.09	2.34E-04	8.25E-03
Prevotellaceae	Prevotella	Otu00128	6	-4.41	1.40	1.63E-03	3.15E-02
ERC severity score as a numerical variable							
Streptococcaceae	Streptococcus	Otu00020	728	0.28	0.09	1.08E-03	4.35E-02
Streptococcaceae	Streptococcus	Otu00061	23	0.35	0.08	2.14E-05	1.74E-03
Number of ERC examinations (statistical model with disease severity as a grouped factor variable)							
Streptococcaceae	Streptococcus	Otu00035	602	1.41	0.22	1.11E-10	2.54E-08
Streptococcaceae	Streptococcus	-	9241	0.39	0.10	1.25E-04	2.63E-03
Number of ERC examinations (statistical model with ERC severity score as a numerical variable)							
Streptococcaceae	Streptococcus	Otu00035	606	0.94	0.23	5.43E-05	1.08E-03
Streptococcaceae	Streptococcus	-	9241	0.46	0.11	1.18E-05	4.96E-04

The two contrasts between the control group and the early disease groups contain all the original significant results, for the reader's convenience. The next five comparisons contain only the results that passed assessment of robustness. Model variables common to all models: taxon ~ sequencing run + IBD status.

5. DISCUSSION

5.1. Parkinson's Disease

5.1.1. Prevotellaceae and Biomarkers

The *Prevotellaceae* family contains four recognised genera, namely *Prevotella*, *Aloprevotella*, *Hallella*, and *Paraprevotella* (Rosenberg 2014). In our dataset, *Prevotella* constituted the vast majority of *Prevotellaceae* reads, and the genus was also statistically significantly reduced in abundance in the PD group when the data was analysed at genus level (unpublished). *Prevotella* have also been found in reduced abundance in autistic children's stool, while *Enterobacteriaceae* were increased (e.g. Kang *et al.* 2013, Krajmalnik-Brown *et al.* 2015). *Prevotella* are gram-negative bacteria that are associated with degradation of complex plant polysaccharides (indigestible by human enzymes) and mucin glycoproteins (Arumugam *et al.* 2011, Wu *et al.* 2011a, Chen *et al.* 2007), the latter being secreted as a component of the gel-like mucus on the surface of the gut mucosa. This mucus serves as a barrier to pathogen breach of the epithelial layer and is part of the *mucosal immune system* (Murphy & Weaver 2016). They are common members of the gut microbiota, and are the dominant component of one of the three originally proposed enterotypes (enterotype 2; Arumugam *et al.* 2011). Given our results, this leads to the prediction that PD prevalence should be lower in this enterotype's population.

Although the genus *Prevotella* does not seem to be a major producer of short-chain fatty acids (SCFAs) in general nor of butyrate in particular (but see Chen *et al.* 2017 and references therein), there is evidence that higher *Prevotella* abundances are associated with communities with high production of various SCFAs due to diet-driven bacterial fermentation of complex plant-derived fibers by various SCFAs-producing taxa (e.g. Ou *et al.* 2013, Chen *et al.* 2017). These communities are also associated with high production of thiamine (vitamin B1) and folate (vitamin B9; Arumugam *et al.* 2011, Ou *et al.* 2013). These vitamins, in turn, are found in reduced levels in PD subjects, and there is evidence that dietary supplements based on these have therapeutic effects in PD (dos Santos *et al.* 2009, Haghdoost-Yazdi *et al.* 2012, Luong & Nguyen 2013). SCFAs have an important role in host energy metabolism (den Besten *et al.* 2013). Butyrate, for example, is the main energy source for the colonocytes that line the colon and has a role in preventing colon cancer (Vernocchi *et al.* 2016). It has also been shown to be an important mediator of colonic inflammatory response by promoting the proliferation of regulatory T cells and increasing the production and circulation of anti-inflammatory cytokines in the gut (Hoeppli *et al.* 2015). The latter is interesting given that pro-inflammatory cytokines and oxidative stress have been causally linked to neuronal death (González *et al.* 2014, Kelly *et al.* 2014). Evidence has been shown that LPS compromises colon permeability, resulting in LPS leakage from the lumen to enteric neuronal tissues, triggering inflammation and promoting neuronal alpha-synuclein overexpression and deposition (Kelly *et al.* 2014). Butyrate could thus have a role in counteracting LPS-driven gut inflammation. Butyrate as well as butyrate-producing gut communities have also been linked, including experimentaly, to the re-establishment of compromised blood-brain barrier functions in germ-free mice (Al-Asmakh & Hedin 2015 and references therein). Pro-inflammatory cytokines and active immune cells can access the brain especially when the blood-brain barrier is compromised, as can be the case in the context of neurodegenerative diseases, including PD (Houser *et al.* 2017 and references therein). Butyrate is also associated with increased gastrointestinal motility in rats, which is impaired in PD (Soret *et al.* 2010). Decreased levels of SCFAs were detected in children with Pervasive Developmental Disorder Not Otherwise Specified (PDD-NOS) and children with Autism Spectrum Disorder

(ASD; De Angelis *et al.* 2013). *Prevotella* abundance, as well as the abundance of other fermenters like *Coprococcus*, were also found decreased in autistic children (Kang *et al.* 2013, Krajmalnik-Brown *et al.* 2015). Interestingly, faecal concentrations of SCFAs have also been shown to be reduced in PD (Unger *et al.* 2016), and producers like *Coprococcus*, *Roseburia*, and the family *Lachnospiraceae* itself have been detected in lower abundances in PD (Keshavarzian *et al.* 2015, Hill-Burns *et al.* 2017). Prevotella has also been estimated to occur in lower abundances in other PD gut studies (Hasegawa *et al.* 2015, Keshavarzian *et al.* 2015, Unger *et al.* 2016).

On the other hand, Sampson *et al.* (2016), using GMO mice overexpressing alpha-synuclein, produced strong evidence that SCFAs produced by gut microbial communities are necessary to substantially develop motor deficits, neuroinflammation, and amyloid pathology in PD mice models. Raising overexpressing mice gnotobiotically without gut bacteria substantially delayed the onset of motor symptoms and reduced gastrointestinal (GI) dysfunction (constipation) compared to overexpressing mice with gut microbiota. Overexpressing gnotobiotic mice also displayed substantially less alpha-synuclein pathology than overexpressing mice with gut microbiota. Young overexpressing mice treated with antibiotics display little motor dysfunction, similarly to overexpressing gnotobiotic mice, and improved GI function. Colonization of the gnotobiotic mice with bacteria from wild-type mice produced motor and GI dysfunction. The authors also observed lower faecal concentrations of SCFAs in gnotobiotic mice and antibiotic-treated mice, as expected. Treating overexpressing gnotobiotic mice with butyrate, propionate, and acetate SCFAs increased microglia activation, neuroinflammation, and amyloid aggregation. SCFAs' treatment also promoted motor deficits and GI dysfunction when comparing to untreated gnotobiotic mice. Administering the anti-inflammatory minocycline to the SCFA-fed overexpressing mice resulted in reduction of amyloid pathology and improved motor symptoms. Faecal transplants from human PD subjects (but not healthy subjects) to overexpressing mice exacerbated the motor symptoms. The alpha-synuclein overexpressing mice phenotype also affected the composition and structure of the gut microbiota after transplantation in a different way from the wild-type mice, implying an interplay between the microbiota and the host's genotype and phenotype. Overall, the study by Sampson *et al.* (2016) suggests that substantial levels of at least some SCFAs are linked with PD neuroinflammation and motor symptoms, in contrast with previous studies showing a reduction of SCFAs in PD (Unger *et al.* 2016) as well as lower levels of *Prevotella* (Hasegawa *et al.* 2015, Keshavarzian *et al.* 2015, Unger *et al.* 2016). However, lower levels of *Prevotella* alone can be compatible with overall high levels of SCFAs, since many other bacterial groups are involved in their production, including *Clostridium* cluster IV (Lopetuso *et al.* 2013, Riviere *et al.* 2016). We detected an increase in abundance in *Clostridiales Incertae Sedis IV*, which could be connected to production of SCFAs.

Decreased mucin production has been associated with increased colonic mucosal permeability, including in PD (Brown *et al.* 2011, Forsyth *et al.* 2011). Increased microbial-induced butyrate production is associated with increased mucin production (Finnie *et al.* 1995, Wright *et al.* 2000, Burger-van Paassen *et al.* 2009). *Prevotella*, being mucin degraders (Wright *et al.* 2000), will tend to be associated with environments rich in mucin, and there is evidence that they are associated with bacterial communities rich in SCFAs' production (e.g. Ou *et al.* 2013, Chen *et al.* 2017), including butyrate, most probably as a consequence of increased fiber consumption in the diet. Assuming these relationships are correct, lower *Prevotella* levels in PD correlate well with reduced mucin production and associated increase in mucosal permeability. Increased mucosal permeability can in turn lead to facilitated entry of toxins and/or pathogens into the host organism, which in turn could be aetiologically associated with PD (Niehaus & Lange 2003). Interestingly, alpha-synuclein was detected in elevated levels in colonic mucosal

staining associated with elevated gut permeability in PD (Forsyth *et al.* 2011), although this study was based on only 9 patients and 10 control subjects, and thus the putative increase in gut permeability in PD needs to be validated. Endotoxin (LPS) administration in mice models showed an increase in gut permeability and alpha-synuclein expression, leading to detection of alpha-synuclein accumulation in colonic myenteric neurons and in the brainstem, mimicking pre-motor pathology in humans (Kelly *et al.* 2014).

Finally, lower levels of ghrelin were found to correlate with lower abundance of *Prevotellaceae* and higher abundance of *Lactobacillaceae* in the colon of rat models subjected to different regimens of nutrition and physical activity, showing that these affect microbial community composition (Queipo-Ortuno *et al.* 2013). Ghrelin is a gut hormone involved in regulation of nigrostriatal dopamine function, and may have a role in the prevention of neurodegeneration in PD (Bayliss *et al.* 2011). Impaired ghrelin secretion has also been shown to be present in PD subjects (Unger *et al.* 2011), and *Prevotellaceae* and *Lactobacillaceae* abundances varied in the same way as in our study (Queipo-Ortuno *et al.* 2013). However, in rats the increase in *Prevotella* abundance was associated with low calory intake, so it is not possible to rule out that these similarities are governed by a preference for certain diet regimens in PD. Feeding restriction (fasting) has also been previously associated with increased abundances of mucin-degraders, including *Prevotella*, which supports a diet-based explanation (Miller & Hoskins 1981, Deplancke *et al.* 2002). Unfortunately, a limitation of our studies is that it was not possible to assess diet data for analysis. To our knowledge, there is no general agreement regarding dietary preferences in PD, although some evidence suggests that PD patients preferentially adopt diets high in vegetable-derived fibers and carbohydrates (Marczewska *et al.* 2006), which would suposedly promote high *Prevotella* levels (Wu *et al.* 2011a, Jeffery & O'Toole 2013). If this is the case, diet alone cannot explain the overall lower relative abundance of gut *Prevotella* in PD. On the other hand, Hill-Burns *et al.* (2017) report that PD patients consume less fruits and vegetables, which supports *Prevotella* abundance being driven by diet preferences. Also, orthostatic hypotension, which affects 30-40% of PD patients, may produce dizziness, visual perturbations, and loss of consciousness, and in the more elderly patients it may tend to occur after feeding (Jost 2003, Iodice *et al.* 2011). This could conceivably lead to changes in diet as an attempt by the patient to ameliorate symptoms.

Overall, it is difficult to envision a mechanism by which lower levels of *Prevotella* could be causally related to Parkinson's disease. Note that although high abundances in our data are never found in PD unlike in controls, low levels are common to both groups, which argues against low *Prevotella* abundance in itself being causally related to PD. Regardless of the explanation for *Prevotella* differential abundance, it is important to view diseases and their causes in terms of the *Sufficient-Component Cause Model* proposed by Rothman (Rothman 1976, Rothman 2012). In this model, also known as the *causal pie model*, a given disease results from one or more *sufficient causes*. These sufficient causes inevitably produce the disease. Any given sufficient cause is composed of *component causes*. These component causes represent factors that, as a group, are all necessary and sufficient to cause a disease. Individually, they don't cause the disease. The interesting part is that any particular disease can have more than one sufficient cause, representing different "pathways" to the disease. As such, this model is of particular interest to microbiome-based biomedical research, since it is unlikely that simplistic cause-effect explanations will elucidate diseases such as Parkinson's or Alzheimer's if indeed these are aetiologically related to the human microbiome. Thus, levels of *Prevotella* below a certain abundance threshold could perhaps be a *necessary component cause* of all sufficient cause models for PD (if more than one pathway exists) for reasons unknown at the moment, but it is not sufficient by itself. Alternatively,

low levels of *Prevotella* could be a component cause in only some sufficient cause pathways. Most probably, such complex diseases as PD involve microbial community composition shifts, with various microbial metabolic pathways involved, that somehow interact with the host's immune system, neurological system, and general physiology. Nevertheless, it is necessary to first elucidate if *Prevotella* abundances are a consequence of dietary effects in PD.

Other possible confounding factors besides diet warrant some discussion. Active or persistent Irritable Bowel Syndrome (IBS) and Inflammatory Bowel Disease (IBD) were exclusion criteria in studies I and II (eTable 1 in supplementary material to Article I). Constipation is associated with PD and there is strong evidence for its effects on gut bacterial community composition (e.g. Zhu *et al.* 2014). The study by Zhu and colleagues detected lower levels of *Prevotella* in the gut of constipated patients, even though the number of subjects was low (8 constipated subjects *vs.* 14 controls; 16S rRNA gene survey study). The cut-off for positive diagnosis of constipation using the Wexner score is 15 (Agachan *et al.* 1996), which we felt was not adequate to meaningfully discriminate between constipated and non-constipated subjects, at least in our study groups. In fact, the highest scoring subject in our dataset (among both PD and control subjects) had a score of 14. Therefore, the Wexner score seems to require a severe degree of constipation to allow diagnosis as such, making a prevalence analysis in our study difficult. On the other hand, the Rome-III questionnaire (Longstreth *et al.* 2006), which was designed to evaluate functional bowel disorders, was deemed inferior for our purposes to the Wexner score, on the basis of the former not being validated through the use of objective physiological measurements (colonic transit time, anal manometry, cinedefecography, electromyography, etc). Thus, we opted to use Wexner as a graded severity score for constipation as opposed to a method for binary diagnosis of the same. This enabled us to use a severity scale during modelling to control for potential confounding effects. In our study, constipation didn't seem to have any substantial effect on *Prevotellaceae* abundance ($p = 0.74$, Table 2 in Article I), with the only significant effect being associated with study group.

Stool consistency has been linked with changes in microbial community composition (Vandeputte *et al.* 2016) making use of the Bristol Stool Scale (BSS) classification system (Lewis & Heaton 1997) and 16S rRNA gene sequencing. Lower levels of *Prevotella* were linked with harder stool and slower transit times, while higher levels of *Ruminococcaceae* were associated with the same. As Vandeputte and colleagues noted, this could be a potential confounder in our study, since the microbial abundance differences observed could reflect stool consistency differences in controls *vs.* PD patients that are related to higher prevalence/severity of constipation in PD and not specifically related to the disease. The authors propose that using the BSS classification would help to disentangle effects from transit time and water activity from effects associated specifically with PD. However, it may be difficult in practice to distinguish between both effects, given that the expected end results on the microbiota could very well be similar, that is, there could be a substantial degree of overlap between the BSS classification and the measured severity of constipation with the Wexner score. Nevertheless, we think their concerns are valid, and transit time and water activity, as measured through the proxy of stool consistency, have some potential to confound results in stool-based studies.

Finally, regarding the potential effects of medications on Study I, we have included warfarin (an anticoagulant) and statins (lipid-lowering medications) in our models. Unfortunately (for us) only two PD subjects were completely free of any Parkinsonian medication, which meant that controlling for the effect of these drugs was not possible in an effective and satisfactory way, even though we detected a possible effect from COMT-inhibitors in the GLMs (Table 2 in Article I). Hill-Burns *et al.* (2017) also detected a statistically significant effect for COMT-

inhibitors and anticholinergics. Importantly, COMT-inhibitors and anticholinergics can produce gastrointestinal side-effects. Thus, we are of the opinion that antiparkinsonian medication has the potential to be a confounder and should, as much as possible, be taken into account during the design of future studies, although this may be difficult to achieve in practice given the "sample of convenience" nature of so many of these studies.

Since the publication of our article on human gut microbiota and PD in 2015, which was the first study on the subject, a further four articles have been published at the time of this writing. Hasegawa *et al.* (2015) found that *Lactobacillus* were present in higher abundance in PD than in controls, which supports our findings. *Prevotella* were estimated as being more abundant in controls than in PD, which also supports our findings, but was not statistically significant. They were also the bacterial group that showed the greatest reduction in abundance in PD in their dataset. However, the study may have lacked statistical power (45 PD patients and 35 controls), and the controls were spouses, which may bias the estimations e.g. due to similar daily diets. The study used RT-PCR of 16S- and 23S rRNA as opposed to 16S rRNA gene amplicon sequencing. Keshavarzian *et al.* (2015) report that *Prevotella* abundances were reduced in sigmoid mucosal biopsies of PD as compared to controls (with less than half the mean PD abundance in controls), but again the result is not statistically significant. Interestingly, the actual stool material showed virtually the same mean abundance in both PD subjects and controls. The number of samples was somewhat low, with 34 controls and 38 PD subjects. Unger *et al.* (2016) report that *Prevotellaceae* were also reduced in PD but not statistically significant. *Lactobacillaceae* were reduced in PD, unlike our results. Their study is based on q-PCR and used 34 PD subjects and a corresponding number of age-matched controls, again on the low side. Finally, Hill-Burns *et al.* (2017) report that *Lactobacillaceae* were increased in abundance in PD, but don't detect any variation in *Prevotellaceae* or *Prevotella*, except for a low abundance *Prevotella* OTU that was increased in PD. *Ruminococcaceae* were positively correlated with PD status. This family in our study (I) was found to be strongly associated only with *Prevotellaceae* abundance levels and not with disease status or any other clinical variable, suggesting that its increase in abundance in PD could be potentially explained either as occupation of available free niche space or due to direct, interspecific competition. Finally, their study was based on 197 cases and 130 controls. However, the study has one main issue: the stool samples were collected using sterile swabs (BD BBL™ CultureSwab™ Sterile, Media-free Swabs from Fisher Scientific), apparently without immersion in any kind of DNA stabilizing solution, and shipped to the destination laboratory through standard United States postal service at ambient temperature. These swabs are apparently manufactured for taking samples and maintaining viability of cells for posterior culturing. However, their suitability to preserve complex community compositional relative abundances, especially for days, is doubtful. According to the authors, the postal service transit time was higher on average for the PD subjects' samples (1-10 days, median 3 days, average 3.2 days) than for the controls' (1-8 days, median 2 days, average 2.6 days). The authors attempt to control for this time discrepancy with statistical modelling, which is reasonable. However, no model-based adjustment is possible for the fact that all the stool samples have been in transit for days in the US postal service, at room temperature and apparently with no DNA stabilizing solution. Thus, the study's results could be compromised to an undetermined extent.

It is interesting to note that all these studies, with the possible exception of Hill-Burns *et al.* (2017), support our results for a substantial *Prevotellaceae*/*Prevotella* abundance reduction in PD, although statistical significance is never achieved even though the estimated effect sizes are large. This could occur due to a variety of reasons, including different criteria for inclusion of cases and controls, different data analysis methods, and use of different clinical variables. Another

contributing factor could be that the number of biological samples included in these studies was relatively low except for Hill-Burns *et al.* (2017), and thus the high inter-individual variation in *Prevotella* abundance could result in a lack of statistical power to achieve statistical significance. Nevertheless, the point estimates were consistently lower for *Prevotellaceae*/*Prevotella* in four out of five studies, including ours (I), and the estimated effect sizes large and similar across the studies. Unfortunately, Hill-Burns *et al.* (2017) do not report the estimates for *Prevotellaceae* or *Prevotella* levels in their study, on the grounds that they were not statistically significant.

Our results suggest that *Prevotella* (or *Prevotellaceae*) abundances could be used as a biomarker, regardless of the existence of an aetiological connection between *Prevotella* and PD. The ROC curve and logistic regression analyses point in this direction. It is clear from the first of these analyses that *Prevotellaceae* abundance is a powerful discriminator for the identification of positive PD cases, but is unable to distinguish the latter efficiently from the healthy subjects, leading to a high percentage of controls being erroneously identified as cases along with the true positives. This could be due to the fact that, although high *Prevotellaceae* abundances are not present in PD, low *Prevotellaceae* abundances are common in both PD and healthy subjects. The addition of other variables reduces the amount of false positives substantially, but comes at the cost of increasing the number of PD cases erroneously identified as negatives. Future studies, like the ones we are now working on, may potentially increase the sensitivity and specificity of the models, and perhaps in the future it will be possible to use a combination of microbiome information from different taxa together with clinical data to develop diagnostic tools for PD that could ideally be used before the onset of motor symptoms (Wu *et al.* 2011b). This is important, given that the reported accuracy of clinical diagnosis for PD varies between 26% and 92%, improving with disease progression (Jankovic *et al.* 2000, Adler *et al.* 2014).

5.1.2. PIGD and Enterobacteriaceae

Enterobacteriaceae abundance was reported as increased in PD by Unger *et al.* (2016). We did not find this differential abundance in our data when contrasting PD patients with controls, but we did find a differential abundance of *Enterobacteriaceae* between the Tremor Dominant (TD) and Postural Instability and Gait Difficulty (PIGD) groups, with the abundance being positively correlated with PIGD severity. In PD, patients with non-tremor dominant phenotype show faster disease progression and more severe alpha-synuclein pathology in the colonic enteric nervous system (Lebouvier *et al.* 2010, Marras & Lang 2013).

Intestinal permeability, for which evidence exists for an increase in PD, has been positively correlated with intestinal mucosal staining for *E. coli* and alpha-synuclein (Forsyth *et al.* 2011), although the sample size in this study is small. LBP levels (LPS-binding protein) were lower in PD subjects, with low levels of LBP having been shown previously to support a role for endotoxin in promoting inflammatory pathways and cytokine production (Forsyth *et al.* 2011). However, this study didn't detect an actual increase in serum LPS levels in PD. Nevertheless, *Escherichia/Shigella* belong to the *Enterobactereaceae* family, which interestingly was found to be positively correlated with the PIGD phenotype in our study.

As mentioned previously, reduced levels of SCFAs are present in PD, and low levels of butyrate and other SCFAs seem to be associated with *Prevotella*-poor gut microbiomes. High fiber-content diets lead to high amounts of SCFAs, which lower the pH of the colon, since the host releases bicarbonate during absorption of SCFAs. This lower pH affects the abundance and composition of the colonic bacteria, which in turn affects the production of SCFAs, potentially leading to even lower SCFAs' levels (den Besten *et al.* 2013). Concentration of SCFAs decreases

from the upper parts of the colon to the rectum, thereby increasing the pH in the same direction (Jacobs & Lupton 1986, Cummings *et al.* 1987, Ward & Coates 1987, Annison *et al.* 2003). An increase in pH in these distal parts of the colon, especially if promoted by abnormally low amounts of SCFAs, can reduce the abundance of butyrate-producing bacteria (or even almost eradicate them) and increase the abundance of *Enterobacteriaceae* (Prohaszka *et al.* 1990, Cherrington *et al.* 1991, Walker *et al.* 2005, Duncan *et al.* 2009, den Besten *et al.* 2013).

However, in what way *Enterobacteriaceae* could potentially influence motor symptoms in PD remains unknown, although endotoxin-induced cascades or neurotransmitters of bacterial origin are a possibility (Niehaus & Lange 2003), as discussed in section 1.2. On the other hand, it is possible that the *Enterobacteriaceae* abundance increase in PIGD is simply reflecting increased pH levels in the colon. However, to our knowledge, no study has shown higher pH levels in PIGD patients compared to TD patients.

5.1.3. Oral and Nasal Bacteria in Parkinson's Disease

The results for alpha and beta diversity with the oral and nasal datasets are supported by previous studies that show higher diversity levels in bacterial oral communities when compared to nasal communities, as well as community compositions that are more similar to each other according to anatomical area than according to specific individuals (Bassis *et al.* 2014). The general community composition of our oral dataset is also in accordance with previously published studies, showing a dominance of *Streptococcus* and a wide distribution of opportunistic pathogens, including in healthy individuals (Bik *et al.* 2010, Palmer 2014).

Prevotella and *Prevotellaceae* were found increased in abundance in the oral samples of PD patients when compared to controls. The genus includes well known opportunistic oral pathogens (Tanaka *et al.* 2008). This led us to speculate that this increase could be related to hygiene practices. Lower frequencies of tooth brushing, less frequent visits for medical dental care, as well as higher prevalence of caries, periodontal disease, and tooth loss have all been reported in PD (Muller *et al.* 2011), but a more recent study reported efficient dental health care among PD patients based on self-assessment (Barbe *et al.* 2017), leaving the question open. Curiously, our data showed that several of the taxa associated with opportunistic oral pathogenicity were actually reduced in PD (seven out of fourteen OTUs, two out of five genera, and one out of five families, all potentially being or including pathogenic opportunistic taxa), which argues against hygiene as the sole explanation for these differential abundances. The other possible explanation, alone or in tandem with hygiene practices, is dietary preference. Some studies suggest that e.g. *Prevotella* grows preferentially on vegetable fibers and carbohydrates, and that PD patients consume more vegetable proteins and carbohydrates (Marczewska *et al.* 2006, Wu *et al.* 2011a, Jeffery & O'Toole 2013), thus potentially explaining the observed abundance difference for this genus in particular. On the other hand, one study reports that diet doesn't significantly affect oral bacterial communities (De Filippis *et al.* 2014). Again, the evidence is inconclusive. All the remaining taxa that were not associated with opportunistic pathogens were found decreased in abundance in PD, with the exception of *Moryella* and *Erysipelotrichaceae*, but further investigation suggests that these two taxa may be false positives resulting from chance effects.

Finally, we found evidence for an increase in putative oral pathogens in male subjects, regardless of PD status. This result is supported by studies showing worse oral hygiene practices among the male population in general (Helldán & Helakorpi 2014). Overall, the existence of differences in oral microbial composition between PD patients and controls suggests that

these could be potentially exploited as biomarkers, regardless of causation, but that potential is nevertheless very limited.

The nasal dataset shows a community composition similar to that reported in previous studies (Bassis *et al.* 2014, Biswas *et al.* 2015). No alpha or beta diversity differences were found between controls and PD patients. As for differentially abundant taxa between these two groups, the genus *Marmoricola*, decreased in PD, has only been isolated from environmental sources, suggesting it may be a lab contaminant (Kim *et al.* 2015). *Flavobacteriaceae* were estimated as being reduced in PD, and this family contains genera known to be human pathogens. However, no OTUs or genera belonging to this family were detected as differentially abundant. Also, its reduced abundance in PD argues against an aetiological connection with PD. Finally, a *Staphylococcus* OTU was found increased in abundance in PD. Oligotyping produced three oligotypes, one resembling *S. epidermis*, another *S. aureus*, and an unresolved *Staphylococcus*. All had similar relative abundances between the PD and control groups, producing no statistically significant hits. In conclusion, our study suggests that nasal microbial communities are not related to Parkinson's disease and offer no biomarker potential.

5.2. Primary Sclerosing Cholangitis

We investigated alpha diversity of the bile communities in control, early disease, advanced disease, and dysplagia/carcinoma groups. Analysis of the samples suggests that diversity is similar in the control group and the early disease stage subjects, and that it then drops progressively throughout the advanced and dysplasia/carcinoma stages, with the latter also showing the sharpest diversity decline. These results are in line with previously reported findings for PSC in the literature (Rossen *et al.* 2015, Kummen *et al.* 2016, Sabino *et al.* 2016), although those studies are based on stool samples as opposed to bile samples. However, given that the source of bile microbiota is thought to originate in the duodenum, these parallel findings are worth considering.

5.2.1. No Aetiological Role for Microbiota?

Except for the *Neisseriaceae* OTU (Otu00213), none of the reported taxa, including those detected at family level analysis, showed credible patterns under visual inspection, suggesting that they could be false positives. Furthermore, the *Enterobacteriaceae*, *Neisseria*, *Campylobacter*, *Neisseriaceae*, and *Clostridiales* OTUs (Otu0008, Otu00045, Otu00089, Otu00213, and Otu00188, respectively) were all found reduced in the early disease stage group, which argues against an etiological role for these groups in PSC. In fact, all OTUs detected were reduced in abundance in the early disease stage, with only families showing a putative increase. Also, except for the *Enterobacteriaceae* OTU, all OTUs show very low to extremely low mean abundances.

The families *Pasteurellaceae*, *Staphylococcaceae*, and *Xanthomonadaceae* contain known human pathogens, but as a group all are quite diverse. Thus, given that detection in all cases occurred only at family level and with no corresponding genera or OTUs detected, the results do not convincingly suggest an infectious role for specific organisms in these families in the initiation of PSC, although the family as a whole could still conceivably influence naïve PSC development. Furthermore, *Xanthomonadaceae* are present in very low mean abundance. The only family detected in common in both early disease models was *Staphylococcaceae*, members of which (e.g. *Staphylococcus aureus*) have been previously detected, e.g. in PSC cultures (Olsson *et al.* 1998) and in sequencing-based studies (Hiramatsu *et al.* 2000).

Although the statistical significance of the above results is suspicious, it would be premature to fully disregard them as false positives, and therefore it is possible that some results may be validated in future studies. Nevertheless, as a whole, our findings fail to provide convincing evidence for the direct involvement of bile microbiota in the aetiology of PSC.

5.2.2. *Streptococcus* and Disease Progression

Both *Streptococcus* as a genus as well as one OTU belonging to this taxon (Otu00020) were found to increase in abundance from early disease to advanced disease stage. From the latter to the dysplasia/carcinoma stage, two different *Streptococcus* OTUs are also seen to increase in abundance (Otu00035 and Otu00061). The same pattern was seen when the analysis was performed with the ERC severity score as a numerical variable, with the abundance of two *Streptococcus* OTUs (Otu00020 and Otu00061, each in common with one of the two previous comparisons) being positively correlated with disease severity. Overall, the results show a generalised streptococcal increase matching disease progression. Notice that this effect is detected even though the number of ERC examinations were used for effect adjustment in the models.

Assessment of the number of ERC examinations in two separate models, in turn, also resulted in the detection of an increase in abundance of the *Streptococcus* genus and one OTU belonging to this group (Otu00035, the same in both models), which is in line with previous studies linking nosocomial infections due to cannulation during ERC, with alpha-haemolytic streptococci being the main culprits (Olsson *et al.* 1998). Hiramatsu *et al.* (2000) detected *Streptococcus milleri* in one out of five PSC patients in gallbladder bile cultures and in sequenced material from the same individual.

However, notice that *Streptococcus* is also present in our control group in similar abundances to the early disease subjects. This could be a consequence of our control subjects maybe not being fully representative of a "healthy" bile, since all had medical indication to undergo ERC due to inconclusive findings during magnetic resonance cholangiopancreatography (MRCP) of the bile ducts or due to elevation of serum alkaline phosphatase (ALP) of unknown origin. The ERC, and in most cases liver histology, excluded the presence of PSC. If the bile of the control subjects is not representative of "normal" bile, then that could also explain why no clear differences were found between controls and naïve PSC subjects. However, this control group is by necessity a sample of convenience, since it would be difficult (and perhaps unethical) to obtain bile samples through ERC from volunteers with no reason to undergo such an intrusive procedure. At present, there is no agreement on what microbial communities in "normal" bile duct-derived bile, if any, should look like (Verdier *et al.* 2015). Overall, our results support the hypothesis that streptococcal infection occurs due to ECR cannulation and is not related to the aetiology of PSC, but also that the nosocomial infection thus developed does seem to have a role in the progressive development of the disease.

6. CONCLUSIONS

The abundances of *Prevotellaceae* and its genus *Prevotella* in the gut (I) were found to be negatively correlated with PD. Given the very high within-group variation in the abundance of these taxonomic groups it is doubtful that this factor alone would lead to PD development. However, the groups' potential as biomarkers for PD is promising, especially combined with other taxa and clinical data, given the need to evaluate the risk of development of PD before amyloid pathology inception. When PD is traditionally diagnosed, it is already too late to apply treatments that could potentially avoid or significantly delay neuronal death. Study limitations regarding confounding by factors like diet and medication remain, but nevertheless the *reduced-Prevotella* hypothesis generated by Study I has subsequently been given support by three independent studies on the microbiome in PD at the time of this writing, with a fourth study regrettably not reporting the estimated effect size and direction for *Prevotella* or *Prevotellaceae* abundance.

Enterobacteriaceae were found to be positively correlated with motor phenotype, specifically postural instability and gait difficulty, suggesting the possibility that this group could somehow affect motor control. The mechanisms, however, remain speculative and were not targeted in our studies. No differences in alpha diversity were found between PD and control groups, suggesting that species richness and evenness are similar among the two populations. On the other hand, beta diversity analysis showed differences at family-level, suggesting that there are significant compositional differences in the gut bacterial communities between the two populations.

As for the oral and nasal microbiota (II), the results for alpha and beta diversity are supported by previous studies that show higher diversity levels in oral communities than nasal communities, as well as community compositions that are more similar to each other according to anatomical area than to specific individuals. The oral communities showed a dominance of *Streptococcus* and a wide distribution of opportunistic pathogens, including in healthy individuals. Curiously, our data suggested that many of the taxa associated with opportunistic pathogens were actually reduced in PD. Overall, the evidence is stronger for differences between PD and controls in the oral area, but no overall clear pattern emerges. Our study suggests that oral communities show little potential for use as biomarkers.

Regarding PSC (III), we found no convincing evidence for a role of bile microbiota in the aetiology of the disease. Differential abundance analysis between controls and naïve PSC patients resulted in the detection of some taxa showing problematic features that suggest they may be false positives, and most were reduced in abundance. Nevertheless, future studies may eventually validate some of those results. More robust were the findings for disease progression, suggesting a clear increase in streptococcal abundances as disease severity increases and finally reaches neoplasia. Our results also support previous studies that suggest that members of the *Streptococcus* genus are associated with nosocomial infections due to cannulation of the bile ducts during ERC. Alpha diversity analysis suggested that diversity decreases as the disease increases in severity, with a sharp diversity drop in the neoplasia group. Limitations regarding the characterization of "healthy bile" and the use of proper controls remain, both in our study and for studies on PSC in general.

7. ACKNOWLEDGEMENTS

It took me a long time to get to the point of writing these acknowledgments. Back in September 2011, when I joined the DNA Sequencing and Genomics Laboratory, the idea was for me to work on the Geomicrobiology of deep-sea hydrothermal vents near the Azores islands, and play to the strengths of my background in Geology. As it turned out, I ended up working exclusively in a variety of projects in the health & disease field, and eventually I chose to make Parkinson's disease and primary sclerosing cholangitis the topic of the thesis you're now holding in your hands (or reading online). That was quite a ride, in no small part due to my academic background in the Earth sciences. I also remember clearly when Petri Auvinen, accepting me as a PhD student of his, told me "you're gonna have to learn Statistics." I think I got an indigestion that day. Ironically, as so many things in life, Biostatistics and Epidemiology became my main academic interests.

Fortunately, I was lucky to have the support of an immense cadre of people to help me out along that journey. I'd like to start by thanking my supervisor Petri Auvinen, who put his faith on a wild card with a background in Geology coming out of nowhere ("I just want to know if you have *the drive* or not"). I wouldn't be defending this thesis if it wasn't for his gamble. Hopefully I didn't disappoint too much. A big thank you & hug to Lars Paulin, who taught me a lot, and with saintly patience answered all my ignorant questions throughout the years (and still does). A warm thanks to Professors Johanna Björkroth (thank you for the shotgun shooting lessons! We need to do that again!) and Liisa Holm, who kindly gave me their support, both as thesis committee advisors as well as outside of that role. Also an extra thanks to Professor Liisa Holm for accepting to be my Custos. Warm thanks to Docent Anne Salonen and Professor Lars Engstrand, who took the time to pre-examine my thesis. I'm not lying when I say that the final version is much better than the original thanks to both of you. A big thanks also to Filip Scheperjans and Martti Färkillä for the opportunity to work with them on the interesting projects that form the basis of this thesis. And thanks to Eero Pekkonen, Tuomas Mertsalmi and Reeta Levo too. This has been an enjoyable collaboration. And of course, warm thanks to everyone who was a co-author in the articles presented in this thesis. And finally, thank you Professor Reetta Satokari for accepting to be my opponent for the thesis defence.

Back at the lab, a special thanks to my long-time collaborator Velma Aho, who has been a co-author in virtually every article and project I've been involved with. Her personal interest in Bioinformatics and programming were quite helpful to me ("Hey, could you write a script to automate this stuff, if I ask nicely?"). Kaisa Koskinen has in the meanwhile moved on to sunnier pastures, but she was extremely helpful as a teacher when I first came to the lab. Thanks! Pia Laine ("Hello! How are you?") has been an endless source of fun moments and actual help (albeit the latter only very occasionally). Olli-Pekka Smolander and Juhana Kammonen have not only been a source of support and a vehicle for interesting nerdy conversations, but have become good friends too. Thanks, guys. I wish you the best of luck. Eeva-Marja Turkki and Kirsi Lipponen have been invaluable in their support, both inside and outside of the wet-lab. My wet-lab adventures would have been a lot more painful if it wasn't for you two. I wish you both the best in life! Same for Harri Kangas, not at our lab anymore, who became a friend and who's help was priceless ("What have you been reading lately?"). Same for Matias Rantanen. Thanks, man! Patrik Koskinen was only for a short time at our lab but that time was enough to find him to be a nice guy. Best of luck! A big thanks to Margarita Andreevskaya, always helpful and ready with a smile. Good luck! Thanks also to Anne Ylinen, Annina Lyyski, Ilhan Cem Duru, Ursula

Lönnqvist, Teija Ojala, Eevakaisa Vesanen, Olli Lounela, Sergei Belanov, Lorenzo Pellegrini, Päivi Laamanen, Hanna Turunen, and Paula Collin-Olkkonen, who have been good lab mates and gave me their support whenever I needed. I truly appreciate it.

Last, but certainly not least, my family and friends. I can't begin to thank the love and support of my father, mother, brother, grandmother Avó Bia, my uncles tio Américo and tia Leonor, and my cousin João. I love you all, very much. This thesis is especially for you, father. The support of my wife, Cátia Suomalainen Pedrosa, has been priceless all these years. Without her I'd never have achieved this. I love you very much, Princess <3! A heartfelt thank you to my father-in-law Armando and mother-in-law Pirjo. I appreciate all the love you've given us. A big hug and kiss to Eva Diamanti and her family. Thank you! A big thanks also to my friends Ana Soares, Olga Korka, Ville and Anna, Päivi, Emmi, Eila and Tommi, JP and Mari, Léo and Marietta, Rita and David, and Joana Revez and Mirko for their support. Finally, a VERY special thanks to my best friends Fernando, Luís, António, Joana Soares, and Nikita Voronov. We went through a lot, guys. Your friendship is a honour. Keep strong and kisses!

Pedro Pereira
Helsinki and Barreiro, August 2017

8. REFERENCES

Adler CH, Beach TG, Hentz JG, Shill HA, Caviness JN, Driver-Dunckley E *et al.* (2014) Low clinical diagnostic accuracy of early vs advanced Parkinson disease: clinicopathologic study. *Neurology,* **83,** 406-412.

Agachan F, Chen T, Pfeifer J, Reissman P, Wexner SD (1996) A constipation scoring system to simplify evaluation and management of constipated patients. *Dis Colon Rectum,* **39,** 681-685.

Aguzzi A, Rajendran L (2009) The transcellular spread of cytosolic amyloids, prions, and prionoids. *Neuron,* **64,** 783-790.

Aho VT, Pereira PA, Haahtela T, Pawankar R, Auvinen P, Koskinen K (2015) The microbiome of the human lower airways: a next generation sequencing perspective. *World Allergy Organ J,* **8,** 23.

Al-Asmakh M, Hedin L (2015) Microbiota and the control of blood-tissue barriers. *Tissue Barriers,* **3,** e1039691.

Anders S, Huber W (2010) Differential expression analysis for sequence count data. *Genome Biol,* **11,** R106.

Anderson MJ (2001) A new method for non-parametric multivariate analysis of variance. *Austral Ecology,* **26,** 32-46.

Anderson MJ, Crist TO, Chase JM, Vellend M, Inouye BD, Freestone AL *et al.* (2011) Navigating the multiple meanings of beta diversity: a roadmap for the practicing ecologist. *Ecol Lett,* **14,** 19-28.

Annison G, Illman RJ, Topping DL (2003) Acetylated, propionylated or butyrylated starches raise large bowel short-chain fatty acids preferentially when fed to rats. *J Nutr,* **133,** 3523-3528.

Arumugam M, Raes J, Pelletier E, Le Paslier D, Yamada T, Mende DR *et al.* (2011) Enterotypes of the human gut microbiome. *Nature,* **473,** 174-180.

Bacci G, Paganin P, Lopez L, Vanni C, Dalmastri C, Cantale C *et al.* (2016) Pyrosequencing Unveils Cystic Fibrosis Lung Microbiome Differences Associated with a Severe Lung Function Decline. *PLoS One,* **11,** e0156807.

Banack SA, Caller TA, Stommel EW (2010) The cyanobacteria derived toxin Beta-N-methylamino-L-alanine and amyotrophic lateral sclerosis. *Toxins (Basel)*, **2**, 2837-2850.

Barbe AG, Bock N, Derman SH, Felsch M, Timmermann L, Noack MJ (2017) Self-assessment of oral health, dental health care and oral health-related quality of life among Parkinson's disease patients. *Gerodontology,* **34,** 135-143.

Barnhart MM, Chapman MR (2006) Curli biogenesis and function. *Annu Rev Microbiol,* **60,** 131-147.

Bassis CM, Tang AL, Young VB, Pynnonen MA (2014) The nasal cavity microbiota of healthy adults. *Microbiome,* **2,** 27.

Bauer TM, Steinbruckner B, Brinkmann FE, Ditzen AK, Schwacha H, Aponte JJ *et al.* (2001) Small intestinal bacterial overgrowth in patients with cirrhosis: prevalence and relation with spontaneous bacterial peritonitis. *Am J Gastroenterol,* **96,** 2962-2967.

Bayliss J, Stark R, Reichenbach A, Andrews ZB (2011) Gut hormones restrict neurodegeneration in Parkinson's disease. *Advanced Understanding of Neurodegenerative Diseases* (ed. R.C.C. Chang), pp. 269-284. Springer, New York.

Beach TG, Adler CH, Sue LI, Vedders L, Lue L, White Iii CL *et al.* (2010) Multi-organ distribution of phosphorylated alpha-synuclein histopathology in subjects with Lewy body disorders. *Acta Neuropathol,* **119,** 689-702.

Beigelman A, Weinstock GM, Bacharier LB (2014) The relationships between environmental bacterial exposure, airway bacterial colonization, and asthma. *Curr Opin Allergy Clin Immunol,* **14,** 137-142.

Bergquist A, Ekbom A, Olsson R, Kornfeldt D, Loof L, Danielsson A *et al.* (2002) Hepatic and extrahepatic malignancies in primary sclerosing cholangitis. *J Hepatol,* **36,** 321-327.

Bested AC, Logan AC, Selhub EM (2013a) Intestinal microbiota, probiotics and mental health: from Metchnikoff to modern advances: Part I - autointoxication revisited. *Gut Pathog,* **5,** 5.

Bested AC, Logan AC, Selhub EM (2013b) Intestinal microbiota, probiotics and mental health: from Metchnikoff to modern advances: Part II - contemporary contextual research. *Gut Pathog,* **5,** 3.

Bested AC, Logan AC, Selhub EM (2013c) Intestinal microbiota, probiotics and mental health: from Metchnikoff to modern advances: part III - convergence toward clinical trials. *Gut Pathog,* **5,** 4.

Bhattacharjee S, Lukiw WJ (2013) Alzheimer's disease and the microbiome. *Front Cell Neurosci,* **7,** 153.

Bianconi E, Piovesan A, Facchin F, Beraudi A, Casadei R, Frabetti F *et al.* (2013) An estimation of the number of cells in the human body. *Ann Hum Biol,* **40,** 463-471.

Bik EM, Long CD, Armitage GC, Loomer P, Emerson J, Mongodin EF *et al.* (2010) Bacterial diversity in the oral cavity of 10 healthy individuals. *ISME J,* **4,** 962-974.

Biswas K, Hoggard M, Jain R, Taylor MW, Douglas RG (2015) The nasal microbiota in health and disease: variation within and between subjects. *Front Microbiol,* **9,** 134.

Blanco LP, Evans ML, Smith DR, Badtke MP, Chapman MR (2012) Diversity, biogenesis and function of microbial amyloids. *Trends Microbiol,* **20,** 66-73.

Bourlat SJ, Borja A, Gilbert J, Taylor MI, Davies N, Weisberg SB *et al.* (2013) Genomics in marine monitoring: new opportunities for assessing marine health status. *Mar Pollut Bull,* **74,** 19-31.

Braak H, de Vos RA, Bohl J, Del Tredici K (2006) Gastric alpha-synuclein immunoreactive inclusions in Meissner's and Auerbach's plexuses in cases staged for Parkinson's disease-related brain pathology. *Neurosci Lett,* **396,** 67-72.

Braak H, Del Tredici K, Rub U, de Vos RA, Jansen Steur EN, Braak E (2003a) Staging of brain pathology related to sporadic Parkinson's disease. *Neurobiol Aging,* **24,** 197-211.

Braak H, Ghebremedhin E, Rub U, Bratzke H, Del Tredici K (2004) Stages in the development of Parkinson's disease-related pathology. *Cell Tissue Res,* **318,** 121-134.

Braak H, Rub U, Gai WP, Del Tredici K (2003b) Idiopathic Parkinson's disease: possible routes by which vulnerable neuronal types may be subject to neuroinvasion by an unknown pathogen. *J Neural Transm (Vienna),* **110,** 517-536.

Bragg L, Stone G, Imelfort M, Hugenholtz P, Tyson GW (2012) Fast, accurate error-correction of amplicon pyrosequences using Acacia. *Nat Methods,* **9,** 425-426.

Brook I (1989) Aerobic and anaerobic microbiology of biliary tract disease. *J Clin Microbiol,* **27,** 2373-2375.

Brown CT, Davis-Richardson AG, Giongo A, Gano KA, Crabb DB, Mukherjee N *et al.* (2011) Gut microbiome metagenomics analysis suggests a functional model for the development of autoimmunity for type 1 diabetes. *PLoS One,* **6,** e25792.

Brundin P, Melki R, Kopito R (2010) Prion-like transmission of protein aggregates in neurodegenerative diseases. *Nat Rev Mol Cell Biol,* **11,** 301-307.

Burger-van Paassen N, Vincent A, Puiman PJ, van der Sluis M, Bouma J, Boehm G *et al.* (2009) The regulation of intestinal mucin MUC2 expression by short-chain fatty acids: implications for epithelial protection. *Biochem J,* **420,** 211-219.

Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK *et al.* (2010) QIIME allows analysis of high-throughput community sequencing data. *Nat Methods,* **7,** 335-336.

Carpenter HA (1998) Bacterial and parasitic cholangitis. *Mayo Clin Proc,* **73,** 473-478.

Chauhan A, Jeans AF (2015) Is Parkinson's disease truly a prion-like disorder? An appraisal of current evidence. *Neurol Res Int,* **2015,** 345285.

Chen J, Wright K, Davis JM, Jeraldo P, Marietta EV, Murray J *et al.* (2016a) An expansion of rare lineage intestinal microbes characterizes rheumatoid arthritis. *Genome Med,* **8,** 43.

Chen SG, Stribinskis V, Rane MJ, Demuth DR, Gozal E, Roberts AM *et al.* (2016b) Exposure to the Functional Bacterial Amyloid Protein Curli Enhances Alpha-Synuclein Aggregation in Aged Fischer 344 Rats and Caenorhabditis elegans. *Sci Rep,* **6,** 34477.

Chen T, Long W, Zhang C, Liu S, Zhao L, Hamaker BR (2017) Fiber-utilizing capacity varies in Prevotella- versus Bacteroides-dominated gut microbiota. *Sci Rep,* **7,** 2594.

Cherrington CA, Hinton M, Pearson GR, Chopra I (1991) Short-chain organic acids at ph 5.0 kill Escherichia coli and Salmonella spp. without causing membrane perturbation. *J Appl Bacteriol,* **70,** 161-165.

Chu Y, Kordower JH (2015) The prion hypothesis of Parkinson's disease. *Curr Neurol Neurosci Rep,* **15,** 28.

Cole JR, Wang Q, Cardenas E, Fish J, Chai B, Farris RJ *et al.* (2009) The Ribosomal Database Project: improved alignments and new tools for rRNA analysis. *Nucleic Acids Res,* **37,** D141-145.

Csendes A, Burdiles P, Maluenda F, Diaz JC, Csendes P, Mitru N (1996) Simultaneous bacteriologic assessment of bile from gallbladder and common bile duct in control subjects and patients with gallstones and common duct stones. *Arch Surg,* **131,** 389-394.

Cumming G (2012) *Understanding the New Statistics: Effect Sizes, Confidence Intervals, and Meta-Analysis*. Routledge.

Cummings JH, Pomare EW, Branch WJ, Naylor CP, Macfarlane GT (1987) Short chain fatty acids in human large intestine, portal, hepatic and venous blood. *Gut,* **28,** 1221-1227.

De Angelis M, Piccolo M, Vannini L, Siragusa S, De Giacomo A, Serrazzanetti DI *et al.* (2013) Fecal microbiota and metabolome of children with autism and pervasive developmental disorder not otherwise specified. *PLoS One,* **8,** e76993.

De Filippis F, Vannini L, La Storia A, Laghi L, Piombino P, Stellato G *et al.* (2014) The same microbiota and a potentially discriminant metabolome in the saliva of omnivore, ovo-lacto-vegetarian and Vegan individuals. *PLoS One,* **9,** e112373.

de Lau LM, Breteler MM (2006) Epidemiology of Parkinson's disease. *Lancet Neurol,* **5,** 525-535.

Debelius J, Song SJ, Vazquez-Baeza Y, Xu ZZ, Gonzalez A, Knight R (2016) Tiny microbes, enormous impacts: what matters in gut microbiome studies? *Genome Biology,* **17**.

Del Tredici K, Braak H (2012) Spinal cord lesions in sporadic Parkinson's disease. *Acta Neuropathol,* **124,** 643-664.

Del Tredici K, Braak H (2016) Review: Sporadic Parkinson's disease: development and distribution of alpha-synuclein pathology. *Neuropathol Appl Neurobiol,* **42,** 33-50.

den Besten G, van Eunen K, Groen AK, Venema K, Reijngoud DJ, Bakker BM (2013) The role of short-chain fatty acids in the interplay between diet, gut microbiota, and host energy metabolism. *J Lipid Res,* **54,** 2325-2340.

Deplancke B, Vidal O, Ganessunker D, Donovan SM, Mackie RI, Gaskins HR (2002) Selective growth of mucolytic bacteria including Clostridium perfringens in a neonatal piglet model of total parenteral nutrition. *Am J Clin Nutr,* **76,** 1117-1125.

Di Rienzi SC, Sharon I, Wrighton KC, Koren O, Hug LA, Thomas BC *et al.* (2013) The human gut and groundwater harbor non-photosynthetic bacteria belonging to a new candidate phylum sibling to Cyanobacteria. *Elife,* **2,** e01102.

Diao J, Burre J, Vivona S, Cipriano DJ, Sharma M, Kyoung M *et al.* (2013) Native alpha-synuclein induces clustering of synaptic-vesicle mimics via binding to phospholipids and synaptobrevin-2/VAMP2. *Elife,* **2,** e00592.

Dickson DW, Fujishiro H, DelleDonne A, Menke J, Ahmed Z, Klos KJ *et al.* (2008) Evidence that incidental Lewy body disease is pre-symptomatic Parkinson's disease. *Acta Neuropathol,* **115,** 437-444.

Dinan TG, Cryan JF (2017b) The Microbiome-Gut-Brain Axis in Health and Disease. *Gastroenterol Clin North Am,* **46,** 77-89.

dos Santos EF, Busanello EN, Miglioranza A, Zanatta A, Barchak AG, Vargas CR *et al.* (2009) Evidence that folic acid deficiency is a major determinant of hyperhomocysteinemia in Parkinson's disease. *Metab Brain Dis,* **24,** 257-269.

Duncan SH, Louis P, Thomson JM, Flint HJ (2009) The role of pH in determining the species composition of the human colonic microbiota. *Environ Microbiol,* **11,** 2112-2122.

Eaton JE, Talwalkar JA, Lazaridis KN, Gores GJ, Lindor KD (2013) Pathogenesis of primary sclerosing cholangitis and advances in diagnosis and management. *Gastroenterology,* **145,** 521-536.

Edgar RC, Haas BJ, Clemente JC, Quince C, Knight R (2011) UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics,* **27,** 2194-2200.

Edwards U, Rogall T, Blocker H, Emde M, Bottger EC (1989) Isolation and direct complete nucleotide determination of entire genes. Characterization of a gene coding for 16S ribosomal RNA. *Nucleic Acids Res,* **17,** 7843-7853.

Eksteen B (2014) Advances and controversies in the pathogenesis and management of primary sclerosing cholangitis. *Br Med Bull,* **110,** 89-98.

Eren AM, Maignien L, Sul WJ, Murphy LG, Grim SL, Morrison HG *et al.* (2013) Oligotyping: Differentiating between closely related microbial taxa using 16S rRNA gene data. *Methods Ecol Evol,* **4**.

Ezkurdia I, Juan D, Rodriguez JM, Frankish A, Diekhans M, Harrow J *et al.* (2014) Multiple evidence strands suggest that there may be as few as 19,000 human protein-coding genes. *Hum Mol Genet,* **23,** 5866-5878.

Fasano A, Bove F, Gabrielli M, Petracca M, Zocco MA, Ragazzoni E *et al.* (2013) The role of small intestinal bacterial overgrowth in Parkinson's disease. *Mov Disord,* **28,** 1241-1249.

Feigelman R, Kahlert CR, Baty F, Rassouli F, Kleiner RL, Kohler P *et al.* (2017) Sputum DNA sequencing in cystic fibrosis: non-invasive access to the lung microbiome and to pathogen details. *Microbiome,* **5,** 20.

Finnie IA, Dwarakanath AD, Taylor BA, Rhodes JM (1995) Colonic mucin synthesis is increased by sodium butyrate. *Gut,* **36,** 93-99.

Foerster CW, Foerster HF (1973) Glutamic acid decarboxylase in spores of Bacillus megaterium and its possible involvement in spore germination. *J Bacteriol,* **114,** 1090-1098.

Folseraas T, Melum E, Rausch P, Juran BD, Ellinghaus E, Shiryaev A *et al.* (2012) Extended analysis of a genome-wide association study in primary sclerosing cholangitis detects multiple novel risk loci. *J Hepatol,* **57,** 366-375.

Forsyth CB, Shannon KM, Kordower JH, Voigt RM, Shaikh M, Jaglin JA *et al.* (2011) Increased intestinal permeability correlates with sigmoid mucosa alpha-synuclein staining and endotoxin exposure markers in early Parkinson's disease. *PLoS One,* **6,** e28032.

Fox J (2016) *Applied Regression Analysis and Generalized Linear Models,* 3rd edn. SAGE Publications.

Fox JG, Yan LL, Dewhirst FE, Paster BJ, Shames B, Murphy JC *et al.* (1995) Helicobacter bilis sp. nov., a novel Helicobacter species isolated from bile, livers, and intestines of aged, inbred mice. *J Clin Microbiol,* **33,** 445-454.

Freedman DA (2009) *Statistical Models: Theory and Practice,* Revised edn. Cambridge University Press.

Ganzle MG, Hertel C, van der Vossen JM, Hammes WP (1999) Effect of bacteriocin-producing lactobacilli on the survival of Escherichia coli and Listeria in a dynamic model of the stomach and the small intestine. *Int J Food Microbiol,* **48,** 21-35.

Garcia M, Lipke P, Klotz S (2013) Pathogenic microbial amyloids: Their function and the host response. *OA Microbiol,* **1**.

Gelman A, Carlin JB, Stern HS, Dunson DB, Vehtari A, Rubin DB (2014) *Bayesian Data Analysis,* 3rd edn. CRC Press, NewYork.

Giesecke J (2002) *Modern Infectious Disease Epidemiology,* 2nd edn. CRC Press.

Goodwin S, McPherson JD, McCombie WR (2016) Coming of age: ten years of next-generation sequencing technologies. *Nat Rev Genet*, **17**, 333-351.

Guthrie GD, Nicholson-Guthrie CS, Leary HL, Jr. (2000) A bacterial high-affinity GABA binding protein: isolation and characterization. *Biochem Biophys Res Commun,* **268,** 65-68.

Haegeman B, Hamelin J, Moriarty J, Neal P, Dushoff J, Weitz JS (2013) Robust estimation of microbial diversity in theory and in practice. *ISME J,* **7,** 1092-1101.

Haghdoost-Yazdi H, Fraidouni N, Faraji A, Jahanihashemi H, Sarookhani M (2012) High intake of folic acid or complex of B vitamins provides anti-Parkinsonism effect: no role for serum level of homocysteine. *Behav Brain Res,* **233,** 375-381.

Hanski I, von Hertzen L, Fyhrquist N, Koskinen K, Torppa K, Laatikainen T *et al.* (2012) Environmental biodiversity, human microbiota, and allergy are interrelated. *Proc Natl Acad Sci U S A,* **109,** 8334-8339.

Harach T, Marungruang N, Duthilleul N, Cheatham V, Mc Coy KD, Frisoni G *et al.* (2017) Reduction of Abeta amyloid pathology in APPPS1 transgenic mice in the absence of gut microbiota. *Sci Rep,* **7,** 41802.

Hasegawa S, Goto S, Tsuji H, Okuno T, Asahara T, Nomoto K *et al.* (2015) Intestinal Dysbiosis and Lowered Serum Lipopolysaccharide-Binding Protein in Parkinson's Disease. *PLoS One,* **10,** e0142164.

Hawkes CH, Del Tredici K, Braak H (2007) Parkinson's disease: a dual-hit hypothesis. *Neuropathol Appl Neurobiol,* **33,** 599-614.

Helldán A, Helakorpi S (2014) *Health Behaviour and Health Among the Finnish Adult Population*. The National Institute for Health and Welfare (THL), Helsinki, Finland.

Hill JM, Lukiw WJ (2015) Microbial-generated amyloids and Alzheimer's disease (AD). *Front Aging Neurosci,* **7,** 9.

Hill-Burns EM, Debelius JW, Morton JT, Wissemann WT, Lewis MR, Wallen ZD *et al.* (2017) Parkinson's disease and Parkinson's disease medications have distinct signatures of the gut microbiome. *Mov Disord*.

Hirai Y (1999) The interaction of bile acids and Helicobacter pylori. *J Gastroenterol,* **34,** 653-654.

Hiramatsu K, Harada K, Tsuneyama K, Sasaki M, Fujita S, Hashimoto T *et al.* (2000) Amplification and sequence analysis of partial bacterial 16S ribosomal RNA gene in gallbladder bile from patients with primary biliary cirrhosis. *J Hepatol,* **33,** 9-18.

Hirschfield GM, Karlsen TH, Lindor KD, Adams DH (2013) Primary sclerosing cholangitis. *Lancet,* **382,** 1587-1599.

Hoeppli RE, Wu D, Cook L, Levings MK (2015) The environment of regulatory T cell biology: cytokines, metabolites, and the microbiome. *Front Immunol,* **6,** 61.

Holmqvist S, Chutna O, Bousset L, Aldrin-Kirk P, Li W, Bjorklund T *et al.* (2014) Direct evidence of Parkinson pathology spread from the gastrointestinal tract to the brain in rats. *Acta Neuropathol,* **128,** 805-820.

Hufnagel DA, Tukel C, Chapman MR (2013) Disease to dirt: the biology of microbial amyloids. *PLoS Pathog,* **9,** e1003740.

Ieno EN, Zuur AF (2015) *Data Exploration and Visualisation with R*. Highland Statistics Ltd., Newburgh, United Kingdom.

Iodice V, Low DA, Vichayanrat E, Mathias CJ (2011) Cardiovascular autonomic dysfunction in MSA and Parkinson's disease: similarities and differences. *J Neurol Sci,* **310,** 133-138.

Jacobs LR, Lupton JR (1986) Relationship between colonic luminal pH, cell proliferation, and colon carcinogenesis in 1,2-dimethylhydrazine treated rats fed high fiber diets. *Cancer Res,* **46,** 1727-1734.

Jankovic J, Rajput AH, McDermott MP, Perl DP (2000) The evolution of diagnosis in early Parkinson disease. Parkinson Study Group. *Arch Neurol,* **57,** 369-372.

Jeffery IB, O'Toole PW (2013) Diet-microbiota interactions and their implications for healthy living. *Nutrients,* **5,** 234-252.

Jost WH (2003) Autonomic dysfunctions in idiopathic Parkinson's disease. *J Neurol,* **250 Suppl 1,** I28-30.

Kang DW, Park JG, Ilhan ZE, Wallstrom G, Labaer J, Adams JB *et al.* (2013) Reduced incidence of Prevotella and other fermenters in intestinal microflora of autistic children. *PLoS One,* **8,** e68322.

Karlsen TH, Boberg KM (2013) Update on primary sclerosing cholangitis. *J Hepatol,* **59,** 571-582.

Kaufmann SH (2008) Immunology's foundation: the 100-year anniversary of the Nobel Prize to Paul Ehrlich and Elie Metchnikoff. *Nat Immunol,* **9,** 705-712.

Kelly LP, Carvey PM, Keshavarzian A, Shannon KM, Shaikh M, Bakay RA *et al.* (2014) Progression of intestinal permeability changes and alpha-synuclein expression in a mouse model of Parkinson's disease. *Mov Disord,* **29,** 999-1009.

Keshavarzian A, Green SJ, Engen PA, Voigt RM, Naqib A, Forsyth CB *et al.* (2015) Colonic bacterial composition in Parkinson's disease. *Mov Disord,* **30,** 1351-1360.

Kim D, Hofstaedter CE, Zhao C, Mattei L, Tanes C, Clarke E *et al.* (2017) Optimizing methods and dodging pitfalls in microbiome research. *Microbiome,* **5,** 52.

Kim SJ, Lim JM, Hamada M, Ahn JH, Weon HY, Suzuki K *et al.* (2015) Marmoricola solisilvae sp. nov. and Marmoricola terrae sp. nov., isolated from soil and emended description of the genus Marmoricola. *Int J Syst Evol Microbiol,* **65,** 1825-1830.

Klein C, Westenberger A (2012) Genetics of Parkinson's disease. *Cold Spring Harb Perspect Med,* **2,** a008888.

Kordower JH (2014) The prion hypothesis of Parkinson's disease: this hot topic just got hotter. *Mov Disord,* **29,** 988.

Koskinen K, Auvinen P, Bjorkroth KJ, Hultman J (2015) Inconsistent Denoising and Clustering Algorithms for Amplicon Sequence Data. *J Comput Biol,* **22,** 743-751.

Kostic AD, Gevers D, Siljander H, Vatanen T, Hyotylainen T, Hamalainen AM *et al.* (2015) The dynamics of the human infant gut microbiome in development and in progression toward type 1 diabetes. *Cell Host Microbe,* **17,** 260-273.

Kozich JJ, Westcott SL, Baxter NT, Highlander SK, Schloss PD (2013) Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform. *Appl Environ Microbiol,* **79,** 5112-5120.

Krajmalnik-Brown R, Lozupone C, Kang DW, Adams JB (2015) Gut bacteria in children with autism spectrum disorders: challenges and promise of studying how a complex community influences a complex disease. *Microb Ecol Health Dis,* **26,** 26914.

Kruschke JK (2015) *Doing Bayesian Data Analysis: A Tutorial with R, JAGS, and Stan,* 2nd edn. Academic Press/Elsevier.

Kummen M, Holm K, Anmarkrud JA, Nygard S, Vesterhus M, Hoivik ML *et al.* (2016) The gut microbial profile in patients with primary sclerosing cholangitis is distinct from patients with ulcerative colitis without biliary disease and healthy controls. *Gut*.

Lane DJ (1991) 16S/23S rRNA sequencing. *Nucleic acid techniques in bacterial systematics* (eds E. Stackebrandt & M. Goodfellow), pp. 115-175. John Wiley and Sons, Chichester, United Kingdom.

Larsen P, Nielsen JL, Dueholm MS, Wetzel R, Otzen D, Nielsen PH (2007) Amyloid adhesins are abundant in natural biofilms. *Environ Microbiol,* **9,** 3077-3090.

Lazarevic V, Gaia N, Emonet S, Girard M, Renzi G, Despres L *et al.* (2014) Challenges in the culture-independent analysis of oral and respiratory samples from intubated patients. *Front Cell Infect Microbiol,* **4,** 65.

Lebouvier T, Neunlist M, Bruley des Varannes S, Coron E, Drouard A, N'Guyen JM *et al.* (2010) Colonic biopsies to assess the neuropathology of Parkinson's disease and its relationship with symptoms. *PLoS One,* **5,** e12728.

Lederberg J, McCray AT (2001) 'Ome Sweet 'Omics—A Genealogical Treasury of Words. *Scientist,* **15(7):8**.

Leek JT, Scharpf RB, Bravo HC, Simcha D, Langmead B, Johnson WE *et al.* (2010) Tackling the widespread and critical impact of batch effects in high-throughput data. *Nat Rev Genet,* **11,** 733-739.

Lerner A, Bagic A (2008) Olfactory pathogenesis of idiopathic Parkinson disease revisited. *Mov Disord,* **23,** 1076-1084.

Lewis SJ, Heaton KW (1997) Stool form scale as a useful guide to intestinal transit time. *Scand J Gastroenterol,* **32,** 920-924.

Li J, Jia H, Cai X, Zhong H, Feng Q, Sunagawa S *et al.* (2014) An integrated catalog of reference genes in the human gut microbiome. *Nat Biotechnol,* **32,** 834-841.

Li JY, Englund E, Holton JL, Soulet D, Hagell P, Lees AJ *et al.* (2008) Lewy bodies in grafted neurons in subjects with Parkinson's disease suggest host-to-graft disease propagation. *Nat Med,* **14,** 501-503.

Lichtman SN, Keku J, Clark RL, Schwab JH, Sartor RB (1991) Biliary tract disease in rats with experimental small bowel bacterial overgrowth. *Hepatology,* **13,** 766-772.

Lichtman SN, Okoruwa EE, Keku J, Schwab JH, Sartor RB (1992) Degradation of endogenous bacterial cell wall polymers by the muralytic enzyme mutanolysin prevents hepatobiliary injury in genetically susceptible rats with experimental intestinal bacterial overgrowth. *J Clin Invest,* **90,** 1313-1322.

Loftus EV, Jr., Sandborn WJ, Lindor KD, Larusso NF (1997) Interactions between chronic liver disease and inflammatory bowel disease. *Inflamm Bowel Dis,* **3,** 288-302.

Longstreth GF, Thompson WG, Chey WD, Houghton LA, Mearin F, Spiller RC (2006) Functional bowel disorders. *Gastroenterology,* **130,** 1480-1491.

Loong TW (2003) Understanding sensitivity and specificity with the right side of the brain. *BMJ,* **327,** 716-719.

Lopetuso LR, Scaldaferri F, Petito V, Gasbarrini A (2013) Commensal Clostridia: leading players in the maintenance of gut homeostasis. *Gut Pathog,* **5,** 23.

Love MI, Huber W, Anders S (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol,* **15,** 550.

Lozupone C, Knight R (2005) UniFrac: a new phylogenetic method for comparing microbial communities. *Appl Environ Microbiol,* **71,** 8228-8235.

Lozupone C, Lladser ME, Knights D, Stombaugh J, Knight R (2011) UniFrac: an effective distance metric for microbial community comparison. *ISME J,* **5,** 169-172.

Lundmark K, Westermark GT, Olsen A, Westermark P (2005) Protein fibrils in nature can enhance amyloid protein A amyloidosis in mice: Cross-seeding as a disease mechanism. *Proc Natl Acad Sci U S A,* **102,** 6098-6102.

Luong KV, Nguyen LT (2013) The beneficial role of thiamine in Parkinson disease. *CNS Neurosci Ther,* **19,** 461-468.

Lynch SV, Pedersen O (2016) The Human Intestinal Microbiome in Health and Disease. *N Engl J Med,* **375,** 2369-2379.

Lyte M, Cryan JF (eds) (2014) *Microbial Endocrinology: The Microbiota-Gut-Brain Axis in Health and Disease*. Springer, New York.

Marczewska A, De Notaris R, Sieri S, Barichella M, Fusconi E, Pezzoli G (2006) Protein intake in Parkinsonian patients using the EPIC food frequency questionnaire. *Mov Disord,* **21,** 1229-1231.

Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA et al. (2005) Genome sequencing in microfabricated high-density picolitre reactors. *Nature*, **437**, 376-380.

Marras C, Lang A (2013) Parkinson's disease subtypes: lost in translation? *J Neurol Neurosurg Psychiatry,* **84,** 409-415.

Mattner J (2016) Impact of Microbes on the Pathogenesis of Primary Biliary Cirrhosis (PBC) and Primary Sclerosing Cholangitis (PSC). *Int J Mol Sci,* **17**.

McArdle BH, Anderson MJ (2001) Fitting multivariate models to community data: A comment on distance-based redundancy analysis. *Ecology,* **82,** 290–297.

McGuigan L, Callaghan M (2015) The evolving dynamics of the microbial community in the cystic fibrosis lung. *Environ Microbiol,* **17,** 16-28.

McMurdie PJ, Holmes S (2013) phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS One,* **8,** e61217.

McMurdie PJ, Holmes S (2014) Waste not, want not: why rarefying microbiome data is inadmissible. *PLoS Comput Biol,* **10,** e1003531.

McNamee R (2005) Regression modelling and other methods to control confounding. *Occup Environ Med,* **62,** 500-506, 472.

Metchnikoff I (1908) *The Prolongation of Life: Optimistic Studies*. Translated by P. Chalmers Mitchell from 1907 original, G.P. Putnam's Sons.

Meyerhans A, Vartanian JP, Wain-Hobson S (1990) DNA recombination during PCR. *Nucleic Acids Res,* **18,** 1687-1691.

Miller RS, Hoskins LC (1981) Mucin degradation in human colon ecosystems. Fecal population densities of mucin-degrading bacteria estimated by a "most probable number" method. *Gastroenterology,* **81,** 759-765.

Miyake Y, Yamamoto K (2013) Role of gut microbiota in liver diseases. *Hepatol Res,* **43,** 139-146.

Morris A, Beck JM, Schloss PD, Campbell TB, Crothers K, Curtis JL *et al.* (2013) Comparison of the respiratory microbiome in healthy nonsmokers and smokers. *Am J Respir Crit Care Med,* **187,** 1067-1075.

Motulsky H (2010) *Intuitive Biostatistics,* 2nd edn. Oxford University Press.

Muller T, Palluch R, Jackowski J (2011) Caries and periodontal disease in patients with Parkinson's disease. *Spec Care Dentist,* **31,** 178-181.

Mulligan VK, Chakrabartty A (2013) Protein misfolding in the late-onset neurodegenerative diseases: common themes and the unique case of amyotrophic lateral sclerosis. *Proteins,* **81,** 1285-1303.

Murphy K, Weaver C (2016) *Janeway's Immunobiology,* 9th edn. Garland Science.

Natale G, Ferrucci M, Lazzeri G, Paparelli A, Fornai F (2011a) Transmission of prions within the gut and towards the central nervous system. *Prion,* **5,** 142-149.

Natale G, Pasquali L, Paparelli A, Fornai F (2011b) Parallel manifestations of neuropathologies in the enteric and central nervous systems. *Neurogastroenterol Motil,* **23,** 1056-1065.

Niehaus I, Lange JH (2003) Endotoxin: is it an environmental factor in the cause of Parkinson's disease? *Occup Environ Med,* **60,** 378.

Niu XL, Liu L, Song ZX, Li Q, Wang ZH, Zhang JL *et al.* (2016) Prevalence of small intestinal bacterial overgrowth in Chinese patients with Parkinson's disease. *J Neural Transm (Vienna),* **123,** 1381-1386.

Oksanen J, Blanchet FG, Friendly M, Kindt R, Legendre P, McGlinn D *et al.* vegan: Community Ecology Package. Accessed 2015. Available from: http://cran.r-project.org/package=vegan

Olanow CW, Brundin P (2013) Parkinson's disease and alpha synuclein: is Parkinson's disease a prion-like disorder? *Mov Disord,* **28,** 31-40.

Olsson R, Bjornsson E, Backman L, Friman S, Hockerstedt K, Kaijser B *et al.* (1998) Bile duct bacterial isolates in primary sclerosing cholangitis: a study of explanted livers. *J Hepatol,* **28,** 426-432.

Ou J, Carbonero F, Zoetendal EG, DeLany JP, Wang M, Newton K *et al.* (2013) Diet, microbiota, and microbial metabolites in colon cancer risk in rural Africans and African Americans. *Am J Clin Nutr,* **98,** 111-120.

Packey CD, Sartor RB (2009) Commensal bacteria, traditional and opportunistic pathogens, dysbiosis and bacterial killing in inflammatory bowel diseases. *Curr Opin Infect Dis,* **22,** 292-301.

Palmer RJ, Jr. (2014) Composition and development of oral bacterial communities. *Periodontol 2000,* **64,** 20-39.

Pezzulo AA, Kelly PH, Nassar BS, Rutland CJ, Gansemer ND, Dohrn CL *et al.* (2013) Abundant DNase I-sensitive bacterial DNA in healthy porcine lungs and its implications for the lung microbiome. *Appl Environ Microbiol,* **79,** 5936-5941.

Polymenidou M, Cleveland DW (2011) The seeds of neurodegeneration: prion-like spreading in ALS. *Cell,* **147,** 498-508.

Pringsheim T, Jette N, Frolkis A, Steeves TD (2014) The prevalence of Parkinson's disease: a systematic review and meta-analysis. *Mov Disord,* **29,** 1583-1590.

Prohaszka L, Jayarao BM, Fabian A, Kovacs S (1990) The role of intestinal volatile fatty acids in the Salmonella shedding of pigs. *Zentralbl Veterinarmed B,* **37,** 570-574.

Prusiner SB (2012) Cell biology. A unifying role for prions in neurodegenerative diseases. *Science,* **336,** 1511-1513.

Qin J, Li R, Raes J, Arumugam M, Burgdorf KS, Manichanh C *et al.* (2010) A human gut microbial gene catalogue established by metagenomic sequencing. *Nature,* **464,** 59-65.

Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P *et al.* (2013) The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res,* **41,** D590-596.

Queipo-Ortuno MI, Seoane LM, Murri M, Pardo M, Gomez-Zumaquero JM, Cardona F *et al.* (2013) Gut microbiota composition in male rat models under different nutritional status and physical activity and its association with serum leptin and ghrelin levels. *PLoS One,* **8,** e65465.

Quince C, Lanzen A, Davenport RJ, Turnbaugh PJ (2011) Removing noise from pyrosequenced amplicons. *BMC Bioinformatics,* **12,** 38.

Reiff C, Kelly D (2010) Inflammatory bowel disease, gut bacteria and probiotic therapy. *Int J Med Microbiol,* **300,** 25-33.

Rideout JR, He Y, Navas-Molina JA, Walters WA, Ursell LK, Gibbons SM *et al.* (2014) Subsampled open-reference clustering creates consistent, comprehensive OTU definitions and scales to billions of sequences. *PeerJ,* **2,** e545.

Riviere A, Selak M, Lantin D, Leroy F, De Vuyst L (2016) Bifidobacteria and Butyrate-Producing Colon Bacteria: Importance and Strategies for Their Stimulation in the Human Gut. *Front Microbiol,* **7,** 979.

Rosebury T (1962) *Microorganisms indigenous to man*. McGraw-Hill, New York.

Rosenberg E (2014) The Family Prevotellaceae. *The Prokaryotes: Other Major Lineages of Bacteria and The Archaea* (eds E. Rosenberg, E.F. DeLong, S. Lory, E. Stackebrandt & F. Thompson), pp. 825-827. Springer Berlin Heidelberg, Berlin, Heidelberg.

Rossen NG, Fuentes S, Boonstra K, D'Haens GR, Heilig HG, Zoetendal EG *et al.* (2015) The mucosa-associated microbiota of PSC patients is characterized by low diversity and low abundance of uncultured Clostridiales II. *J Crohns Colitis,* **9,** 342-348.

Rothberg JM, Leamon JH (2008) The development and impact of 454 sequencing. *Nat Biotechnol*, **26**, 1117-1124.

Rothman KJ (1976) Causes. *Am J Epidemiol,* **104,** 587-592.

Rothman KJ (2012) *Epidemiology: An Introduction,* 2nd edn. Oxford University Press.

Sabino J, Vieira-Silva S, Machiels K, Joossens M, Falony G, Ballet V *et al.* (2016) Primary sclerosing cholangitis is characterised by intestinal dysbiosis independent from IBD. *Gut,* **65,** 1681-1689.

Salter SJ, Cox MJ, Turek EM, Calus ST, Cookson WO, Moffatt MF *et al.* (2014) Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. *BMC Biol,* **12,** 87.

Sampson TR, Debelius JW, Thron T, Janssen S, Shastri GG, Ilhan ZE *et al.* (2016) Gut Microbiota Regulate Motor Deficits and Neuroinflammation in a Model of Parkinson's Disease. *Cell,* **167,** 1469-1480 e1412.

Sasatomi K, Noguchi K, Sakisaka S, Sata M, Tanikawa K (1998) Abnormal accumulation of endotoxin in biliary epithelial cells in primary biliary cirrhosis and primary sclerosing cholangitis. *J Hepatol,* **29,** 409-416.

Savage DC (1977) Microbial ecology of the gastrointestinal tract. *Annu Rev Microbiol,* **31,** 107-133.

Scher JU, Sczesnak A, Longman RS, Segata N, Ubeda C, Bielski C *et al.* (2013) Expansion of intestinal Prevotella copri correlates with enhanced susceptibility to arthritis. *Elife,* **2,** e01202.

Schloss PD. Standard Operating Procedure for 454 sequenced 16S rRNA gene data. Accessed July, 2013. Available from: http://www.mothur.org/wiki/454_SOP

Schloss PD. Standard Operating Procedure for MiSeq sequenced 16S rRNA gene data. Accessed June and October, 2014. Available from: http://www.mothur.org/wiki/MiSeq_SOP

Schloss PD (2009) A high-throughput DNA sequence aligner for microbial ecology studies. *PLoS One,* **4,** e8230.

Schloss PD (2010) The effects of alignment quality, distance calculation method, sequence filtering, and region on the analysis of 16S rRNA gene-based studies. *PLoS Comput Biol,* **6,** e1000844.

Schloss PD (2013) Secondary structure improves OTU assignments of 16S rRNA gene sequences. *ISME J,* **7,** 457-460.

Schloss PD, Gevers D, Westcott SL (2011) Reducing the effects of PCR amplification and sequencing artifacts on 16S rRNA-based studies. *PLoS One,* **6,** e27310.

Schloss PD, Westcott SL (2011) Assessing and improving methods used in operational taxonomic unit-based approaches for 16S rRNA gene sequence analysis. *Appl Environ Microbiol,* **77,** 3219-3226.

Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB *et al.* (2009) Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl Environ Microbiol,* **75,** 7537-7541.

Schrag A, Horsfall L, Walters K, Noyce A, Petersen I (2015) Prediagnostic presentations of Parkinson's disease in primary care: a case-control study. *Lancet Neurol,* **14,** 57-64.

Schwartz K, Boles BR (2013) Microbial amyloids--functions and interactions within the host. *Curr Opin Microbiol,* **16,** 93-99.

Segal LN, Alekseyenko AV, Clemente JC, Kulkarni R, Wu B, Gao Z *et al.* (2013) Enrichment of lung microbiome with supraglottic taxa is associated with increased pulmonary inflammation. *Microbiome,* **1,** 19.

Sengoku R, Saito Y, Ikemura M, Hatsuta H, Sakiyama Y, Kanemaru K *et al.* (2008) Incidence and extent of Lewy body-related alpha-synucleinopathy in aging human olfactory bulb. *J Neuropathol Exp Neurol,* **67,** 1072-1083.

Shannon KM, Keshavarzian A, Mutlu E, Dodiya HB, Daian D, Jaglin JA *et al.* (2012) Alpha-synuclein in colonic submucosa in early untreated Parkinson's disease. *Mov Disord,* **27,** 709-715.

Shen H, Ye F, Xie L, Yang J, Li Z, Xu P *et al.* (2015) Metagenomic sequencing of bile from gallstone patients to identify different microbial community patterns and novel biliary bacteria. *Sci Rep,* **5,** 17450.

Shreiner AB, Kao JY, Young VB (2015) The gut microbiome in health and in disease. *Curr Opin Gastroenterol,* **31,** 69-75.

Smyth RP, Schlub TE, Grimm A, Venturi V, Chopra A, Mallal S *et al.* (2010) Reducing chimera formation during PCR amplification to ensure accurate genotyping. *Gene,* **469,** 45-51.

Son JS, Zheng LJ, Rowehl LM, Tian X, Zhang Y, Zhu W *et al.* (2015) Comparison of Fecal Microbiota in Children with Autism Spectrum Disorders and Neurotypical Siblings in the Simons Simplex Collection. *PLoS One,* **10,** e0137725.

Soo RM, Hemp J, Parks DH, Fischer WW, Hugenholtz P (2017) On the origins of oxygenic photosynthesis and aerobic respiration in Cyanobacteria. *Science*, **355**, 1436-1440.

Soo RM, Skennerton CT, Sekiguchi Y, Imelfort M, Paech SJ, Dennis PG *et al.* (2014) An expanded genomic representation of the phylum cyanobacteria. *Genome Biol Evol,* **6,** 1031-1045.

Soret R, Chevalier J, De Coppet P, Poupeau G, Derkinderen P, Segain JP *et al.* (2010) Short-chain fatty acids regulate the enteric neurons and control gastrointestinal motility in rats. *Gastroenterology,* **138,** 1772-1782.

Storey JD, Tibshirani R (2003) Statistical significance for genomewide studies. *Proc Natl Acad Sci U S A,* **100,** 9440-9445.

Su MS, Schlicht S, Ganzle MG (2011) Contribution of glutamate decarboxylase in Lactobacillus reuteri to acid resistance and persistence in sourdough fermentation. *Microb Cell Fact,* **10 Suppl 1,** S8.

Surette MG (2014) The cystic fibrosis lung microbiome. *Ann Am Thorac Soc,* **11 Suppl 1,** S61-65.

Sveinbjornsdottir S (2016) The clinical symptoms of Parkinson's disease. *J Neurochem,* **139 Suppl 1,** 318-324.

Sze MA, Hogg JC, Sin DD (2014) Bacterial microbiome of lungs in COPD. *Int J Chron Obstruct Pulmon Dis,* **9,** 229-238.

Tabibian JH, O'Hara SP, Lindor KD (2014) Primary sclerosing cholangitis and the microbiota: current knowledge and perspectives on etiopathogenesis and emerging therapies. *Scand J Gastroenterol,* **49,** 901-908.

Tabibian JH, Talwalkar JA, Lindor KD (2013) Role of the microbiota and antibiotics in primary sclerosing cholangitis. *Biomed Res Int,* **2013,** 389537.

Tanaka S, Yoshida M, Murakami Y, Ogiwara T, Shoji M, Kobayashi S *et al.* (2008) The relationship of Prevotella intermedia, Prevotella nigrescens and Prevotella melaninogenica in the supragingival plaque of children, caries and oral malodor. *J Clin Pediatr Dent,* **32,** 195-200.

R Core Team. R: A language and environment for statistical computing. Available from: https://www.r-project.org/

Torres J, Bao X, Goel A, Colombel JF, Pekow J, Jabri B *et al.* (2016) The features of mucosa-associated microbiota in primary sclerosing cholangitis. *Aliment Pharmacol Ther,* **43,** 790-801.

Torres V (2014) Ten Times More Microbial Cells than Body Cells in Humans? *Microbe Magazine,* **9(2)**.

Tsuneyama K, Harada K, Kono N, Hiramatsu K, Zen Y, Sudo Y *et al.* (2001) Scavenger cells with gram-positive bacterial lipoteichoic acid infiltrate around the damaged interlobular bile ducts of primary biliary cirrhosis. *J Hepatol,* **35,** 156-163.

Turnbaugh PJ, Ley RE, Hamady M, Fraser-Liggett CM, Knight R, Gordon JI (2007) The human microbiome project. *Nature,* **449,** 804-810.

Ulusoy A, Phillips RJ, Helwig M, Klinkenberg M, Powley TL, Di Monte DA (2017) Brain-to-stomach transfer of alpha-synuclein via vagal preganglionic projections. *Acta Neuropathol,* **133,** 381-393.

Unger MM, Moller JC, Mankel K, Eggert KM, Bohne K, Bodden M *et al.* (2011) Postprandial ghrelin response is reduced in patients with Parkinson's disease and idiopathic REM sleep behaviour disorder: a peripheral biomarker for early Parkinson's disease? *J Neurol,* **258,** 982-990.

Unger MM, Spiegel J, Dillmann KU, Grundmann D, Philippeit H, Burmann J *et al.* (2016) Short chain fatty acids and gut microbiota differ between patients with Parkinson's disease and age-matched controls. *Parkinsonism Relat Disord,* **32,** 66-72.

Vandeputte D, Falony G, Vieira-Silva S, Tito RY, Joossens M, Raes J (2016) Stool consistency is strongly associated with gut microbiota richness and composition, enterotypes and bacterial growth rates. *Gut,* **65,** 57-62.

Verdier J, Luedde T, Sellge G (2015) Biliary Mucosal Barrier and Microbiome. *Viszeralmedizin,* **31,** 156-161.

Vernocchi P, Del Chierico F, Putignani L (2016) Gut Microbiota Profiling: Metabolomics Based Approach to Unravel Compounds Affecting Human Health. *Front Microbiol,* **7,** 1144.

Visanji NP, Brooks PL, Hazrati LN, Lang AE (2013) The prion hypothesis in Parkinson's disease: Braak to the future. *Acta Neuropathol Commun,* **1,** 2.

Wakabayashi K, Takahashi H, Takeda S, Ohama E, Ikuta F (1988) Parkinson's disease: the presence of Lewy bodies in Auerbach's and Meissner's plexuses. *Acta Neuropathol,* **76,** 217-221.

Walker AW, Duncan SH, McWilliam Leitch EC, Child MW, Flint HJ (2005) pH and peptide supply can radically alter bacterial populations and short-chain fatty acid ratios within microbial communities from the human colon. *Appl Environ Microbiol,* **71,** 3692-3700.

Ward FW, Coates ME (1987) Gastrointestinal pH measurement in rats: influence of the microbial flora, diet and fasting. *Lab Anim,* **21,** 216-222.

Weiss S, Xu ZZ, Peddada S, Amir A, Bittinger K, Gonzalez A *et al.* (2017) Normalization and microbial differential abundance strategies depend upon data characteristics. *Microbiome,* **5,** 27.

Westcott SL, Schloss PD (2015) De novo clustering methods outperform reference-based methods for assigning 16S rRNA gene sequences to operational taxonomic units. *PeerJ,* **3,** e1487.

White JR, Nagarajan N, Pop M (2009) Statistical methods for detecting differentially abundant features in clinical metagenomic samples. *PLoS Comput Biol,* **5,** e1000352.

Whittaker RH (1960) Vegetation of the Siskiyou Mountains, Oregon and California. *Ecological Monographs,* **30,** 279-338.

Wirdefeldt K, Adami HO, Cole P, Trichopoulos D, Mandel J (2011) Epidemiology and etiology of Parkinson's disease: a review of the evidence. *Eur J Epidemiol,* **26 Suppl 1,** S1-58.

Wright DP, Knight CG, Parkar SG, Christie DL, Roberton AM (2000) Cloning of a mucin-desulfating sulfatase gene from Prevotella strain RS2 and its expression using a Bacteroides recombinant system. *J Bacteriol,* **182,** 3002-3007.

Wu CT, Davis PA, Luketic VA, Gershwin ME (2004) A review of the physiological and immunological functions of biliary epithelial cells: targets for primary biliary cirrhosis, primary sclerosing cholangitis and drug-induced ductopenias. *Clin Dev Immunol,* **11,** 205-213.

Wu GD, Chen J, Hoffmann C, Bittinger K, Chen YY, Keilbaugh SA *et al.* (2011a) Linking long-term dietary patterns with gut microbial enterotypes. *Science,* **334,** 105-108.

Wu Y, Le W, Jankovic J (2011b) Preclinical biomarkers of Parkinson disease. *Arch Neurol,* **68,** 22-30.

Ye F, Shen H, Li Z, Meng F, Li L, Yang J *et al.* (2016) Influence of the Biliary System on Biliary Bacteria Revealed by Bacterial Communities of the Human Biliary and Upper Digestive Tracts. *PLoS One,* **11,** e0150519.

Young VB (2017) The role of the microbiome in human health and disease: an introduction for clinicians. *BMJ,* **356,** j831.

Zhu L, Liu W, Alkhouri R, Baker RD, Bard JE, Quigley EM *et al.* (2014) Structural changes in the gut microbiome of constipated patients. *Physiol Genomics,* **46,** 679-686.

Recent Publications in this Series

28/2017 Anna-Riia Holmström
Learning from Medication Errors in Healthcare — How to Make Medication Error Reporting Systems Work?
29/2017 Aaro Haapaniemi
Laryngeal Cancer Recurrence, Prognostic Factors and Management
30/2017 Reeta Talka
Effects of Opioids on Neuronal Nicotinic Acetylcholine Receptors
31/2017 Andreas Helfenstein
Fighting Bugs by Numbers: Bioinformatics Tools for Antimicrobial Drug Discovery
32/2017 Feven Tigistu-Sahle
Metabolism of Polyunsaturated Fatty Acids in Human Bone Marrow Derived Mesenchymal Stromal Cells
33/2017 Maija Miettinen
Vitamin D and Type 1 Diabetes
34/2017 Antti-Pekka Rissanen
Oxygen Delivery and Utilization during Acute Dynamic Exercise: Effects of Polycystic Ovary Syndrome, Type 1 Diabetes, and Exercise Training
35/2017 Harri Nurmi
Role of the VEGFC/VEGFR3 Pathway, Beyond Developmental Lymphangiogenesis
36/2017 Mohamed Mostafa Mohamed Ahmed El Missiry
Early Treatment Prediction and Immunological Effects of Tyrosine Kinase Inhibitor Therapy in Chronic-Phase Chronic Myeloid Leukemia
37/2017 Salla Keskitalo
CMGC Kinases and Cancer
38/2017 Mikhail Yuryev
In Vivo Imaging of the Early Embryonic Cortex in Rodents
39/2017 Mari Teesalu
Uncovering a Sugar Tolerance Network: SIK3 and Cabut as Downstream Effectors of Mondo-Mlx
40/2017 Katriina Tarkiainen
Pharmacogenetics of Carboxylesterase 1
41/2017 Noora Sjöstedt
In Vitro Evaluation of the Pharmacokinetic Effects of BCRP Interactions
42/2017 Jenni Hällfors
Nicotine Dependence — Identifying the Contribution of Specific Genes
43/2017 Marjaana Pussila
Cancer-preceding Gene Expression Changes in Mouse Colon Mucosa
44/2017 Ansku Holstila
Changes in Leisure-Time Physical Activity, Functioning, Work Disability and Retirement: A Follow-Up Study among Employees
45/2017 Jelena Meinilä
Diet Quality and Its Association with Gestational Diabetes
46/2017 Martina B. Lorey
Secretome Analysis of Human Macrophages Activated by Microbial Stimuli
47/2017 Eeva Suvikas-Peltonen
Lääkkeiden turvallisen käyttökuntoon saattamisen edistäminen sairaaloiden osastoilla

Helsinki 2017 ISSN 2342-3161 ISBN 978-951-51-3607-7